



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## **Evasão do curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília**

Daniel Garcia da Costa

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Orientadora

Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Emilia Machado Telles Walter

Coorientador

Prof. Dr. Jan Mendonça Correa

Brasília  
2017



Evasão do curso de Licenciatura em Matemática  
(Diurno) da Universidade de Brasília

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Prof. Dr. Jan Mendonça Corrêa      Dr.<sup>a</sup> Maria Inez Machado Telles Walter  
CIC/UnB      DPO/UnB

Brasília, 02 de fevereiro de 2017

# Dedicatória

Dedico esse trabalho aos meus pais, pelo apoio incondicional.

Aos professores pela luz do conhecimento.

À Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Emília e ao Prof. Dr. Jan Mendonça, que me deram seus votos de confiança e a oportunidade para que eu possa realizar esse trabalho.

# Agradecimentos

Eu agradeço especialmente aos meus pais que sempre me apoiaram.

Agradeço à Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Emília e ao Prof. Dr. Jan Mendonça, pela paciência e suporte na orientação desse trabalho.

# Resumo

A evasão é um problema que representa um grande desafio para as Instituições de Ensino Superior no Brasil, pois possui diversas causas e provoca graves consequências no âmbito social, acadêmico e econômico. Atualmente existem poucos estudos sobre o problema de evasão do curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília e não há um entendimento claro sobre o perfil de alunos com risco de evasão. Esse trabalho teve como objetivo principal analisar estatisticamente as reprovações em disciplinas obrigatórias do curso e identificar quais dessas disciplinas estão relacionadas à evasão do curso.

**Palavras-chave:** evasão, mineração de dados, análise estatística, perfil de alunos com risco de evasão, Licenciatura em Matemática Diurno, Universidade de Brasília

# Abstract

The dropout rate is a problem that represents a major challenge for the higher education institutions in Brazil, since it is originated from many reasons and causes extensive damage in the social, academic and economic areas. Currently there are few studies on the problem of dropout rate in Mathematics for Teaching degree course at the University of Brasilia while there is no clear understanding of the profile of students. This work aimed to statistically analyze the failures in compulsory subjects of the course, as well as identify the compulsory disciplines that are related to the dropout rate cause in this course.

**Keywords:** dropout rate, data mining, students with avoidance profile, Teaching degree in mathematics, statistical analysis, University of Brasília

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Problema . . . . .	2
1.2	Justificativa . . . . .	2
1.3	Objetivo . . . . .	3
1.4	Descrição dos Capítulos . . . . .	3
<b>2</b>	<b>A Evasão no Ensino Superior</b>	<b>4</b>
2.1	Conceitos Básicos . . . . .	4
2.2	Evasão no Contexto Brasileiro . . . . .	8
2.3	Evasão em Nível Internacional . . . . .	12
2.4	Evasão nas Licenciaturas . . . . .	14
2.5	Evasão na Universidade de Brasília . . . . .	16
<b>3</b>	<b>Mineração de Dados</b>	<b>19</b>
3.1	Dado, Informação e Conhecimento . . . . .	19
3.2	Definições de Mineração de Dados . . . . .	20
3.3	Processos de Extração de Conhecimento em Dados . . . . .	21
3.4	Técnica de Mineração de Dados . . . . .	24
3.5	WEKA . . . . .	27
<b>4</b>	<b>Informações do Curso e dos Alunos</b>	<b>29</b>
4.1	Licenciatura em Matemática (Diurno) . . . . .	29
4.1.1	Descrição do Curso . . . . .	29
4.1.2	Disciplinas Obrigatórias . . . . .	30
4.1.3	O Fluxo Vigente da Habilitação Diurna . . . . .	32
4.2	Dados dos Alunos . . . . .	33
4.2.1	Dados Brutos . . . . .	33
4.2.2	Filtragem e Limpeza dos Dados Brutos . . . . .	35

<b>5</b>	<b>Análises Estatísticas</b>	<b>37</b>
5.1	A Situação Atual . . . . .	37
5.2	Formas de Ingresso no Curso . . . . .	38
5.3	A Saída do Curso . . . . .	40
5.4	Taxas de Reprovação em Disciplinas Obrigatórias . . . . .	41
5.4.1	Primeiro Semestre . . . . .	42
5.4.2	Segundo Semestre . . . . .	45
5.4.3	Terceiro Semestre . . . . .	51
5.4.4	Quarto Semestre . . . . .	56
5.4.5	Quinto Semestre . . . . .	57
5.4.6	Sexto Semestre . . . . .	58
5.4.7	Sétimo Semestre . . . . .	59
5.4.8	Oitavo Semestre . . . . .	59
5.5	Classificação das Disciplinas por Reprovação . . . . .	61
<b>6</b>	<b>Disciplinas Obrigatórias Relacionadas à Evasão</b>	<b>62</b>
6.1	Mineração de Dados . . . . .	62
6.2	Análise dos Resultados . . . . .	67
<b>7</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>71</b>
7.1	Contribuições . . . . .	71
7.2	Trabalhos Futuros . . . . .	72
	<b>Referências</b>	<b>73</b>
	<b>Anexo</b>	<b>77</b>
<b>I</b>	<b>Formato Completo do Banco de Dados</b>	<b>78</b>
<b>II</b>	<b>Currículo Vigente em 1º/2005</b>	<b>81</b>
<b>III</b>	<b>Algoritmo Utilizado para Limpeza dos Dados</b>	<b>82</b>
<b>IV</b>	<b>Algoritmo Utilizado para Inserção dos Dados no MySQL</b>	<b>87</b>



# Lista de Figuras

2.1	Percentual do número de IES e Percentual de Matrículas por Organização Acadêmica. Fonte: INEP [15]. . . . .	9
2.2	Evolução do Número de Ingressos nas IES, por Grau Acadêmico dos Cursos de Graduação. Fonte: INEP [15]. . . . .	11
2.3	Evolução do Número de Concluintes nas IES, por Grau Acadêmico dos Cursos de Graduação. Fonte: INEP [15]. . . . .	11
2.4	A saída no curso de Matemática por sistema de entrada no período de 2º/2004 a 2º/2012. Fonte: DEG/UnB [14]. . . . .	17
3.1	Uma visão geral das etapas que compõem o processo de descoberta de conhecimento em banco de dados ( <i>Knowledge Discovery in Databases/KDD</i> ) (Adaptado de Fayyad et al. [20]). . . . .	22
3.2	Uma visão geral das etapas que compõem o processo CRISP (Adaptado de Chapman et al. [9]). . . . .	23
3.3	Uma visão geral das etapas que compõem o processo SEMMA. (Adaptado de Olson e Delen [39]). . . . .	24
3.4	Primeira etapa de classificação de acordo com Han e Kamber [27]. . . . .	25
3.5	Segunda etapa de classificação de acordo com Han e Kamber [27]. . . . .	26
3.6	Interface principal WEKA - Escolha de aplicações. . . . .	27
3.7	Interface <i>Explorer</i> do WEKA. . . . .	28
4.1	Fluxo de disciplinas do curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília [17]. . . . .	32
5.1	Situação atual dos alunos analisados (2º/1977 a 1º/2016). . . . .	37
5.2	Quantidade de alunos ingressados por forma de ingresso (2º/1977 a 1º/2016). . . . .	38
5.3	Percentual de alunos ingressados pelo sistema universal e por cotas (2º/1977 a 1º/2016). . . . .	39
5.4	Gênero dos alunos ingressados (2º/1977 a 1º/2016). . . . .	39
5.5	Tipo de escola de origem (2º/1977 a 1º/2016). . . . .	39

5.6	Porcentagens de alunos relativas à forma de saída do curso (2º/1977 a 1º/2016).	40
5.7	A saída do curso por sistema de entrada (2º/1977 a 1º/2016).	41
5.8	Taxa de reprovação em Cálculo 1 no período de 2º/1977 a 2º/1997.	42
5.9	Taxa de reprovação em Cálculo 1 no período de 1º/1998 a 1º/2016.	43
5.10	Média de reprovação em Cálculo 1 entre o período de 2º/1977 a 1º/2016.	43
5.11	Taxa de reprovação em Introdução à Ciência da Computação no período de 2º/1977 a 2º/1997.	44
5.12	Taxa de reprovação em Introdução à Ciência da Computação no período de 1º/1998 a 1º/2016.	44
5.13	Média de reprovação em Introdução à Ciência da Computação entre o período de 2º/1977 a 1º/2016.	45
5.14	Taxa de reprovação em Cálculo 2 no período de 2º/1977 a 2º/1997.	45
5.15	Taxa de reprovação em Cálculo 2 no período de 1º/1998 a 1º/2016.	46
5.16	Média de reprovação em Cálculo 2 entre o período de 2º/1977 a 1º/2016.	46
5.17	Taxa de reprovação em Introdução à Álgebra Linear no período de 2º/1977 a 2º/1997.	47
5.18	Taxa de reprovação em Introdução à Álgebra Linear no período de 1º/1998 a 1º/2016.	47
5.19	Média de reprovação em Introdução à Álgebra Linear entre o período de 2º/1977 a 1º/2016.	48
5.20	Taxa de reprovação em Física 1 no período de 2º/1977 a 2º/1997.	48
5.21	Taxa de reprovação em Física 1 no período de 1º/1998 a 1º/2016.	49
5.22	Média de reprovação em Física 1 entre o período de 2º/1977 a 1º/2016.	49
5.23	Taxa de reprovação em Física 1 Experimental no período de 2º/1977 a 2º/1997.	50
5.24	Taxa de reprovação em Física 1 Experimental no período de 1º/1998 a 1º/2016.	50
5.25	Média de reprovação em Física 1 Experimental entre o período de 2º/1977 a 1º/2016.	51
5.26	Taxa de reprovação em Cálculo 3 no período de 2º/1977 a 2º/1997.	51
5.27	Taxa de reprovação em Cálculo 3 no período de 1º/1998 a 1º/2016.	52
5.28	Média de reprovação em Cálculo 3 entre o período de 2º/1977 a 1º/2016.	52
5.29	Taxa de reprovação em Teoria dos Números no período de 2º/1977 a 2º/1997.	53
5.30	Taxa de reprovação em Teoria dos Números no período de 1º/1998 a 1º/2016.	53
5.31	Média de reprovação em Teoria dos Números entre o período de 2º/1977 a 1º/2016.	54

5.32	Taxa de reprovação em Física 2 no período de 2º/1977 a 2º/1997. . . . .	54
5.33	Taxa de reprovação em Física 2 no período de 1º/1998 a 1º/2016. . . . .	55
5.34	Média de reprovação em Física 2 entre o período de 2º/1977 a 1º/2016. . .	55
5.35	Taxa de reprovação em Álgebra 1 no período de 2º/1977 a 2º/1997. . . . .	56
5.36	Taxa de reprovação em Álgebra 1 no período de 1º/1998 a 1º/2016. . . . .	57
5.37	Média de reprovação em Álgebra 1 entre o período de 2º/1977 a 1º/2016. .	57
5.38	Taxa de reprovação em Análise 1 no período de 2º/1977 a 2º/1997. . . . .	58
5.39	Taxa de reprovação em Análise 1 no período de 1º/1998 a 1º/2016. . . . .	59
5.40	Média de reprovação em Análise 1 entre o período de 2º/1977 a 1º/2016. .	59
5.41	Taxa de reprovação em Variável Complexa 1 no período de 2º/1977 a 2º/1997.	60
5.42	Taxa de reprovação em Variável Complexa 1 no período de 1º/1998 a 1º/2016.	60
5.43	Média de reprovação em Variável Complexa 1 entre o período de 2º/1977 a 1º/2016. . . . .	61
6.1	Resultado obtido na mineração de dados com o algoritmo J48, sem o atributo MATRICALUNO. . . . .	63
6.2	Resultado obtido na mineração de dados com o algoritmo J48, sem os atributos MATRICALUNO e FOR SAIDA OPCAIO. . . . .	64
6.3	Resultado obtido na mineração de dados com o algoritmo J48, sem os atributos MATRICALUNO, FOR SAIDA OPCAIO e IRA. . . . .	65
6.4	Atributos considerados na mineração de dados com o algoritmo J48, após a remoção dos atributos adicionais. . . . .	66
6.5	Resultado do algoritmo J48, ao utilizar os atributos mostrados na Figura 6.4.	67
6.6	Média de alunos que reprovaram em Álgebra 1 e evadiram do curso, por menção. . . . .	68
6.7	Alunos que reprovaram em Introdução à Álgebra Linear e evadiram do curso, após terem sido aprovados em Álgebra 1 com menção MM. . . . .	68
6.8	Alunos que reprovaram em Cálculo 3 e evadiram do curso, após terem sido aprovados em Álgebra 1 e Introdução à Álgebra Linear, com menção MM.	69
6.9	Alunos que reprovaram em Cálculo 3 e evadiram do curso, após terem sido aprovados em Álgebra 1 e Introdução à Álgebra Linear, com menção MM.	69
6.10	Alunos que reprovaram em Introdução à Álgebra Linear e evadiram do curso, após terem sido aprovados em Álgebra 1 com menção MS. . . . .	69

# Lista de Tabelas

2.1	Definições de evasão segundo o estudo de Lee e Choi . . . . .	5
2.2	Número de Instituições de Ensino Superior, por Categoria Administrativa. Fonte: INEP [15]. . . . .	9
2.3	Evolução do Número de Cursos de Graduação, por Grau Acadêmico. Fonte: INEP [15]. . . . .	10
2.4	Evolução do Número de Matrículas de Graduação, por Grau Acadêmico. Fonte: INEP [15]. . . . .	10
2.5	Percentual de Matrículas, Ingressantes e Concluintes por Área de Atuação em 2013. Fonte: INEP [15]. . . . .	12
2.6	Estudos realizados sobre a evasão nos cursos de Licenciatura no Brasil. . .	15
2.7	Evasão nos cursos de Ciências Exatas da Universidade de Brasília. Período de 2º/2004 a 2º/2012. . . . .	16
3.1	Dados, informação e conhecimento, segundo Davenport [13]. . . . .	20
4.1	Licenciatura em Matemática (Diurno) - Informações sobre a habilitação de acordo com o Matricula Web [17]. . . . .	30
4.2	Currículo da Licenciatura em Matemática - Disciplinas obrigatórias de acordo com o Matricula Web [17]. . . . .	31
4.3	Os atributos presentes nos dados brutos . . . . .	33
5.1	Classificação das disciplinas que os alunos mais reprovaram. . . . .	61
6.1	Disciplinas obrigatórias que mais estão relacionadas à evasão no curso. . .	70
II.1	Currículo vigente em 1º/2005. . . . .	81

# Capítulo 1

## Introdução

A Lei de Diretrizes e Bases da Educação Nacional (LDBEN) [1], Lei nº 9.394 de 1996, estabelece no artigo 43 a finalidade da Educação Superior pelos seguintes incisos:

- I estimular a criação cultural e o desenvolvimento do espírito científico e do pensamento reflexivo;
- II formar diplomados nas diferentes áreas de conhecimento, aptos para a inserção em setores profissionais e para a participação no desenvolvimento da sociedade brasileira, e colaborar na sua formação contínua;
- III incentivar o trabalho de pesquisa e investigação científica, visando o desenvolvimento da ciência e da tecnologia e da criação e difusão da cultura, e, desse modo, desenvolver o entendimento do homem e do meio em que vive;
- IV promover a divulgação de conhecimentos culturais, científicos e técnicos que constituem patrimônio da humanidade e comunicar o saber através do ensino, de publicações ou de outras formas de comunicação;
- V suscitar o desejo permanente de aperfeiçoamento cultural e profissional e possibilitar a correspondente concretização, integrando os conhecimentos que vão sendo adquiridos numa estrutura intelectual sistematizadora do conhecimento de cada geração;
- VI estimular o conhecimento dos problemas do mundo presente, em particular os nacionais e regionais, prestar serviços especializados à comunidade e estabelecer com esta uma relação de reciprocidade;
- VII promover a extensão, aberta à participação da população, visando à difusão das conquistas e benefícios resultantes da criação cultural e da pesquisa científica e tecnológica geradas na instituição.

Tais incisos reforçam a importância do sistema de Educação Superior para um país desenvolvido e independente. A produção de conhecimento, o desenvolvimento científico e tecnológico são imprescindíveis para o desenvolvimento socioeconômico e cultural do país, pois fomentam a evolução do indivíduo e da sociedade.

O artigo 3º da LDBEN [1] determina que o ensino deve ser ministrado com base nos seguintes princípios: igualdade de condições para o acesso e permanência na escola; e garantia de padrão de qualidade do ensino. O artigo 4º reforça que é dever do Estado com a educação escolar pública garantir padrões mínimos de qualidade de ensino e insumos indispensáveis ao desenvolvimento do processo de ensino-aprendizagem, além de acesso aos níveis mais elevados do ensino, da pesquisa e da criação artística, segundo a capacidade de cada um. Com base nessa lei é possível concluir que o ensino superior está regido pelos princípios de qualidade de ensino.

Um dos grandes problemas que afetam a qualidade do ensino superior é a evasão de alunos. Baggi e Lopes [4] enfatizam que a evasão é um problema que possui diversas origens e vem preocupando as instituições de ensino em geral, públicas ou particulares, e que a evasão de alunos provoca graves consequências sociais, acadêmicas e econômicas.

Os resultados do Censo da Educação Superior de 2013 [15] sobre o ingresso e a evasão no Ensino Superior, indicam que há um percentual muito baixo de alunos formados nos cursos da área de Exatas, tais como Ciências, Matemática e Computação, em relação aos alunos que ingressaram.

Nesse contexto, atualmente, não se tem uma visão clara do problema de evasão do curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília (UnB).

A obtenção dos fatores que causam a evasão de alunos no curso podem colaborar com o Departamento de Matemática/UnB para elaboração de políticas de combate à evasão.

## 1.1 Problema

Não existe estudo detalhado sobre a evasão de alunos no curso de Licenciatura em Matemática (Diurno) da UnB.

## 1.2 Justificativa

Ao não se ter uma visão clara do problema de evasão do curso, pode-se criar políticas de combate à evasão pouco eficientes e eficazes. Assim, um estudo que permita avaliar os fatores de evasão e identificar as características dos alunos em risco pode fornecer informações relevantes para os departamentos que ofertam as disciplinas do curso. Em particular, se as reprovações em um conjunto de disciplinas estiverem ligadas à evasão dos

alunos do Departamento de Matemática da UnB, pode permitir a criação de medidas a fim de diminuir os índices de evasão no curso.

## 1.3 Objetivo

O objetivo geral deste trabalho é fazer análises estatísticas sobre a evasão que afeta o curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília e propor um perfil para alunos em risco de evasão, a partir de um conjunto de disciplinas em que os alunos foram reprovados.

Os objetivos específicos são:

- Realizar análises estatísticas sobre a reprovação em disciplinas obrigatórias;
- Identificar as disciplinas que afetam diretamente na evasão, por mineração de dados;
- Criar um perfil de aluno em risco de evasão, com base nas análises anteriores.

## 1.4 Descrição dos Capítulos

Esse trabalho está dividido em capítulos. No Capítulo 2, estão as definições, fatores e conceitos gerais sobre o problema de evasão no ensino superior. A evasão é descrita em nível internacional, no contexto brasileiro e nos cursos de licenciatura no Brasil.

O Capítulo 3 explica os conceitos gerais sobre mineração de dados, processos e técnicas. Introduz conceitos gerais sobre dado, informação e conhecimento. Ainda, apresenta a ferramenta WEKA, que será usada nesse trabalho.

O Capítulo 4 apresenta as informações básicas sobre o curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília, quais sejam, o currículo, as disciplinas obrigatórias e o fluxo. Ainda, mostra análises dos dados brutos dos alunos.

O Capítulo 5 discute estatísticas de reprovação no curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília, assim como conjuntos de disciplinas com reprovação, que podem estar ligadas à evasão.

O Capítulo 6 apresenta os resultados obtidos pela mineração de dados e as disciplinas que estão ligadas à evasão.

Por fim, o Capítulo 7 conclui este trabalho e indica novas direções de investigação.

# Capítulo 2

## A Evasão no Ensino Superior

Este capítulo apresenta conceitos e informações sobre a evasão no ensino superior. A Seção 2.1 apresenta a importância de estudos sobre a evasão e os diferentes conceitos sobre evasão. A Seção 2.2 apresenta o contexto e as estatísticas sobre a evasão no ensino superior brasileiro. A Seção 2.3 aborda a evasão em nível internacional. Na Seção 2.4, é apresentada a evasão nos cursos de Licenciatura no Brasil. Por fim, a Seção 2.5 mostra a evasão no contexto da Universidade de Brasília.

### 2.1 Conceitos Básicos

A evasão no ensino superior é um grande problema, presente não só nas Instituições de Ensino Superior (IES) do Brasil, mas em todo mundo. Silva Filho et al. [49] afirma que a evasão estudantil no ensino superior é um problema internacional que afeta o resultado dos sistemas educacionais: As graves consequências causadas pela evasão são desperdícios sociais, acadêmicos e econômicos, tanto no setor público quanto no privado. No setor público, a evasão causa a perda de recursos públicos investidos na educação. No setor privado, representa uma importante perda de receitas da instituição. No campo social, a evasão contribui com a falta de profissionais especializados em determinadas áreas de conhecimento. No campo acadêmico, prejudica a avaliação do curso e da instituição.

Para Lobo [34], a evasão representa uma perda social, de recursos e de tempo de todos os envolvidos no processo de ensino: por parte do aluno, dos professores, da instituição de ensino, do sistema educacional e toda a sociedade.

Para entender o motivo da evasão do aluno no ensino superior, é preciso verificar se a evasão está relacionada aos fatores pessoais do aluno, ou da instituição, ou do sistema de ensino. Entretanto essa é uma tarefa complexa, pois exige uma investigação ampla e profunda dos contextos internos e externos no campo acadêmico, social e pessoal.



Lima e Costa [32] afirmam que, devido à complexidade do fenômeno de evasão, torna-se difícil tanto a mensuração quanto a generalização dos resultados dos estudos sobre esse tema. Para começar, a evasão possui diversas abordagens e diferentes conceitos, o que pode causar resultados completamente diferentes para o mesmo estudo de caso.

No estudo realizado por Lee e Choi [30] é possível verificar as diferentes definições de evasão na visão de cada autor:

Tabela 2.1: Definições de evasão segundo o estudo de Lee e Choi [30]

<b>Autores</b>	<b>Ano</b>	<b>Definição</b>
Shin e Kim [48]	1999	Evasão: estudantes que não se matricularam em nenhuma matéria por três semestres seguidos
Morgan e Tam [37]	1999	Não Conclusão: estudantes que não se matricularam para o semestre seguinte
Woodley et al. [55]	2001	Evasão: estudantes que não cumpriram algum dos requisitos do curso, que abandonaram o curso em qualquer momento dele, além daqueles estudantes que infringiram alguma regra da universidade
Xenos et al. [57]	2002	Evasão: estudantes que se matricularam mas não começaram os seus estudos e não se matricularam no ano seguinte, estudantes que cursaram por algum tempo a universidade, mas que em algum momento resolveram desistir
Kemp [29]	2002	Não Conclusão: estudantes que abandonaram antes do início do curso, que se desligaram durante o curso, ou ainda que não conseguiram notas suficientes para passar
Cheung e Kan [10]	2002	Evasão: estudantes não conseguiram notas suficientes para passar
Moore et al. [36]	2003	Não Conclusão: estudantes que não conseguiram notas suficientes para passar, ou que se desligaram oficialmente do curso

Dupin-Bryant [19]	2004	Evasão: estudante que não completou nenhuma matéria durante um semestre
Castles [8]	2004	Evasão: estudante que se desligou formalmente, abandonou sem avisar a universidade, ou que não completou nenhuma matéria durante um semestre
Pierrakeas et al. [43]	2004	Evasão: estudantes que não conseguiram passar em matérias mas comentaram que gostariam de continuar ou alunos que não conseguiram passar em matérias e indicaram que não gostariam de continuar
Willging e Johnson [53]	2004	Evasão: estudantes que abandonaram o curso depois do primeiro semestre
Morris et al. [38]	2005	Desligamento: estudantes que se desligaram oficialmente. Não Conclusão: estudantes que não conseguiram notas suficientes para passar ou não tiveram rendimento
Frydenberg [24]	2007	Evasão: estudantes que se matricularam mas desistiram antes do curso começar ou no início do curso
Levy [31]	2007	Evasão: estudantes que se desligaram voluntariamente por motivos financeiros
Ivankova e Stick [28]	2007	Evasão: estudantes que se desligaram ou foram desligados do curso
Tello [50]	2007	Desistência: estudantes que se desligaram oficialmente do curso antes da sua conclusão
Pigliapoco e Bogliolo [44]	2008	Evasão: estudantes que não se rematricularam depois do primeiro ano de curso

Perry [42]	2008	Desligamento: estudantes que não cumpriram os requerimentos mínimos do curso e estudantes que abandonaram a universidade sem motivos claros
Finnegan et al. [22]	2009	Desligamento: estudantes tiveram que se retirar do curso oficialmente; estudantes que não conseguiram notas suficientes para passar ou não tiveram rendimento

---

No sentido geral, a evasão no ensino superior pode ser caracterizada pela interrupção temporária ou permanente dos estudos, de forma voluntária ou involuntária, influenciada por uma conjunção de fatores internos ou externos ao aluno.

A Comissão Especial de Estudos sobre a Evasão nas Universidades Brasileiras [16] define a evasão no ensino superior em três tipos:

1. Evasão do curso: desligamento do curso superior em função de abandono (não-matrícula), transferência ou reescolha, trancamento e/ou exclusão por norma institucional;
2. Evasão da instituição: desligamento da instituição na qual está matriculado;
3. Evasão do sistema: abandono definitivo ou temporário do ensino superior.

Miranda [35] define a evasão em duas classes, por duração e amplitude:

- Duração:
  - Temporária: Trancamento por obrigação ou voluntário, quando se tem a intenção de voltar.
  - Definitiva: Quando existe abandono, desistência ou transferência externa;
- Amplitude:
  - Do curso.
  - Da instituição;
  - Do sistema educacional;

Lobo [34] reforça o conceito de evasão ao indicar quatro tipos: a evasão do curso, a evasão da instituição de ensino, a evasão do sistema de ensino e a evasão do aluno, que origina todas as outras.

- Do curso:
  - É o tipo de evasão em que o aluno deixa o curso, podendo ou não continuar na mesma IES.
- Da IES:
  - Trata-se da evasão em que o aluno muda de IES, mas não abandona o sistema de ensino.
- Do sistema de ensino:
  - É aquela em que o aluno abandona o sistema de ensino e deixa de estudar em qualquer IES.
- Do aluno:
  - É a evasão que origina todas as outras e só é possível medi-la com precisão por meio do acompanhamento da evolução da situação individualizada de cada estudante.

Baseando-se nessas definições, utilizaremos neste trabalho a definição de evasão **definitiva** (quanto à duração) e do **curso** (quanto à amplitude). Entre os tipos de evasão observados, iremos categorizar o conceito de evasão como **Evasão de Curso**. Para o estudo da evasão, serão consideradas as taxas de reprovação nas disciplinas obrigatórias do curso.

## 2.2 Evasão no Contexto Brasileiro

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)<sup>1</sup> é uma autarquia federal vinculada ao Ministério da Educação (MEC)<sup>2</sup>. Tem como responsabilidade promover pesquisas, estudos, e avaliações sobre o Sistema Educacional Brasileiro. Possui como objetivo produzir informações claras e confiáveis na área de educação, para os pesquisadores, gestores, educadores e público em geral.

O Censo da Educação Superior no Brasil [15], realizado pelo INEP, consiste em levantar e analisar dados estatísticos sobre os cursos, vagas, inscrições, matrículas, ingressantes e concluintes nas Instituições de Ensino Superior (IES). Os dados são coletados a partir do preenchimento dos questionários, por parte das IES e por importação de dados do Sistema e-MEC, e então, é feito um resumo técnico, que disponibiliza informações detalhadas para a comunidade acadêmica e à sociedade em geral.

---

<sup>1</sup>[inep.gov.br](http://inep.gov.br)

<sup>2</sup>[mec.gov.br](http://mec.gov.br)

O último Resumo Técnico do Censo da Educação Superior no Brasil foi realizado em 2013. No total, 2.391 Instituições de Ensino Superior participaram do Censo. As instituições privadas representaram 87,4% do total, enquanto as instituições públicas, apenas 12,6%. A Tabela 2.2 mostra o total de IES por categoria administrativa, classificadas por Federal, Estadual, Municipal e Privada.

Tabela 2.2: Número de Instituições de Ensino Superior, por Categoria Administrativa. Fonte: INEP [15].

Ano	Total	Federal	Estadual	Municipal	Privada
2010	2.378	99	108	71	2.100
2011	2.365	103	110	71	2.081
2012	2.416	103	116	85	2.112
2013	2.391	106	119	76	2.090

A Figura 2.1 mostra a distribuição das IES e o percentual de matrículas por organização acadêmica. Embora 84,3% das IES sejam faculdades, elas concentram apenas 29,2% do total de matrículas. No entanto, as universidades, que representam 8,2% das IES, concentram 53,4% das matrículas.

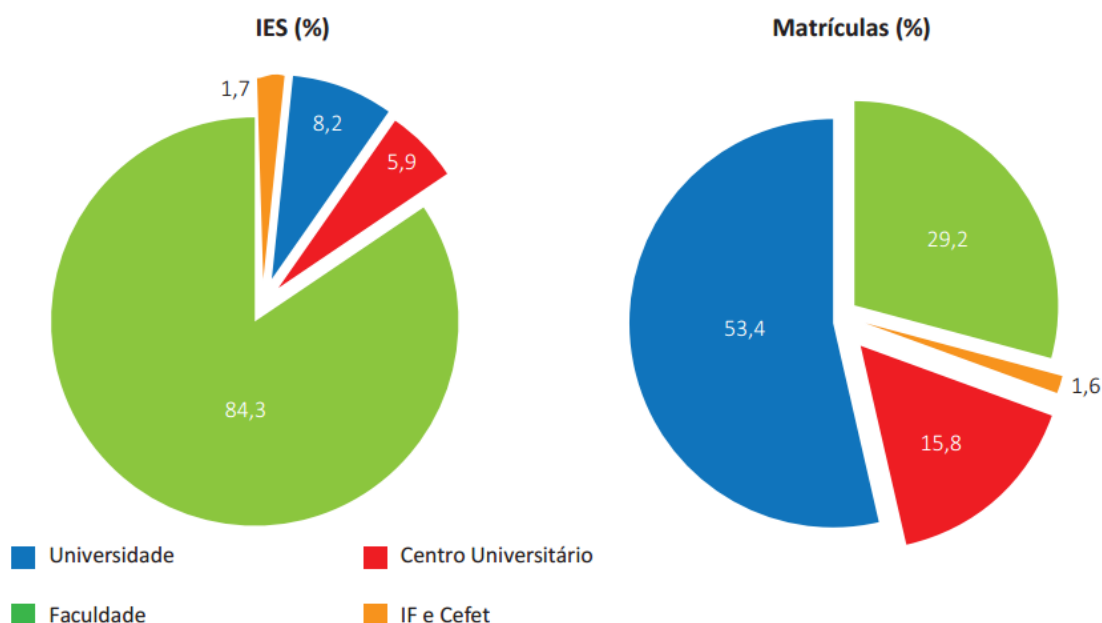


Figura 2.1: Percentual do número de IES e Percentual de Matrículas por Organização Acadêmica. Fonte: INEP [15].

A Tabela 2.3 mostra a evolução do número de cursos de graduação nas IES, por grau acadêmico. É notável que entre os anos de 2012 e 2013, os cursos de Bacharelado aumen-

taram em 1,14%, os cursos Tecnológicos em 4,27%, enquanto os cursos de Licenciatura tiveram uma redução de 3,43%.

Tabela 2.3: Evolução do Número de Cursos de Graduação, por Grau Acadêmico. Fonte: INEP [15].

<b>Ano</b>	<b>Total Geral</b>	<b>Bacharelado</b>	<b>Licenciatura</b>	<b>Tecnológico</b>
2010	29.507	16.586	7.922	4.999
2011	30.420	17.031	7.911	5.478
2012	31.866	17.703	8.194	5.969
2013	32.049	17.905	7.920	6.224

A Tabela 2.4 mostra a evolução do número de matrículas nas IES, por grau acadêmico. Segundo o relatório técnico, no período de 2010 a 2013, os cursos de bacharelado continuaram com o maior número de matrículas, seguidos pelos cursos de Licenciatura. Os cursos tecnológicos obtiveram um crescimento de 27,4% na quantidade de matrículas.

Tabela 2.4: Evolução do Número de Matrículas de Graduação, por Grau Acadêmico. Fonte: INEP [15].

<b>Grau Acadêmico</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
Total	6.379.299	6.739.689	7.037.688	7.305.977
Bacharelado	4.226.717	4.495.831	4.703.693	4.912.310
Licenciatura	1.354.989	1.356.329	1.366.559	1.374.174
Tecnológico	781.609	870.534	944.904	995.746
Não Aplicável	15.984	16.995	22.532	23.747

A Figura 2.2 ilustra a evolução do número de ingressos por grau acadêmico nos cursos de graduação das IES. É notável que, de 2010 a 2011, a quantidade de ingressos nos cursos de Licenciatura é maior que dos cursos tecnológicos. Nos anos de 2012 e 2013, a quantidade de ingressos nos cursos tecnológicos superou a dos cursos de Licenciatura. No período de 2010 a 2013, houve um aumento na taxa de ingressos nos cursos de bacharelado em 29,7%, nos cursos de Licenciatura em 3,7% e nos cursos tecnológicos em 36,6%.

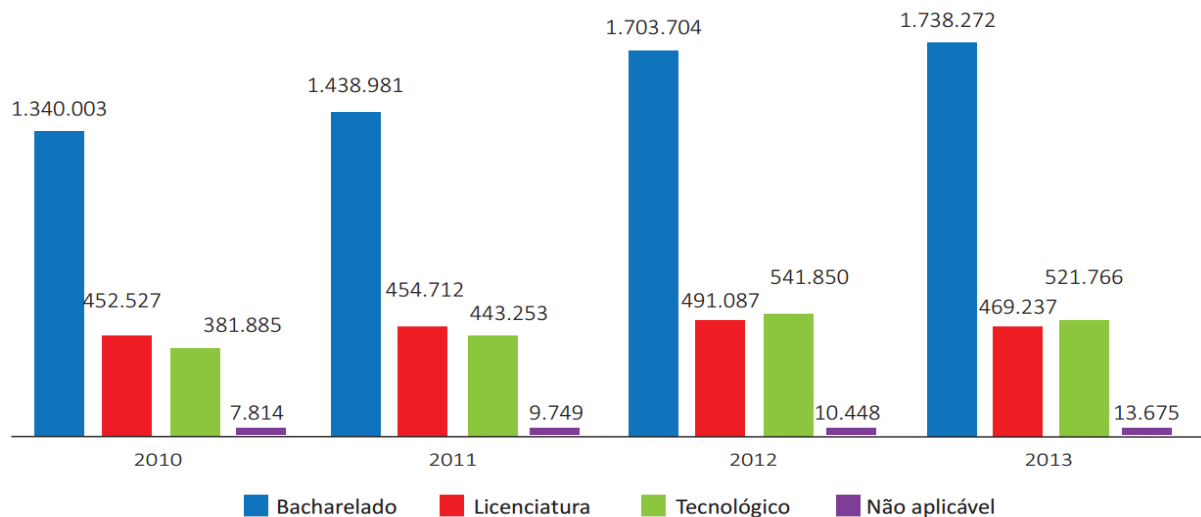


Figura 2.2: Evolução do Número de Ingressos nas IES, por Grau Acadêmico dos Cursos de Graduação. Fonte: INEP [15].

A Figura 2.3 ilustra a evolução do número de concluintes por grau acadêmico nos cursos de graduação das IES. Em 2013, dos 991.010 concluintes, 60,1% formaram-se em cursos de bacharelado, 20,3% em cursos de Licenciatura e 19,7% em cursos tecnológicos. Pode-se observar que, entre 2012 e 2013, houve um aumento dos concluintes para o grau tecnológico (3,1%) e uma queda de concluintes nas Licenciaturas (-10,1%) e nos cursos de bacharelado (-6,7%).

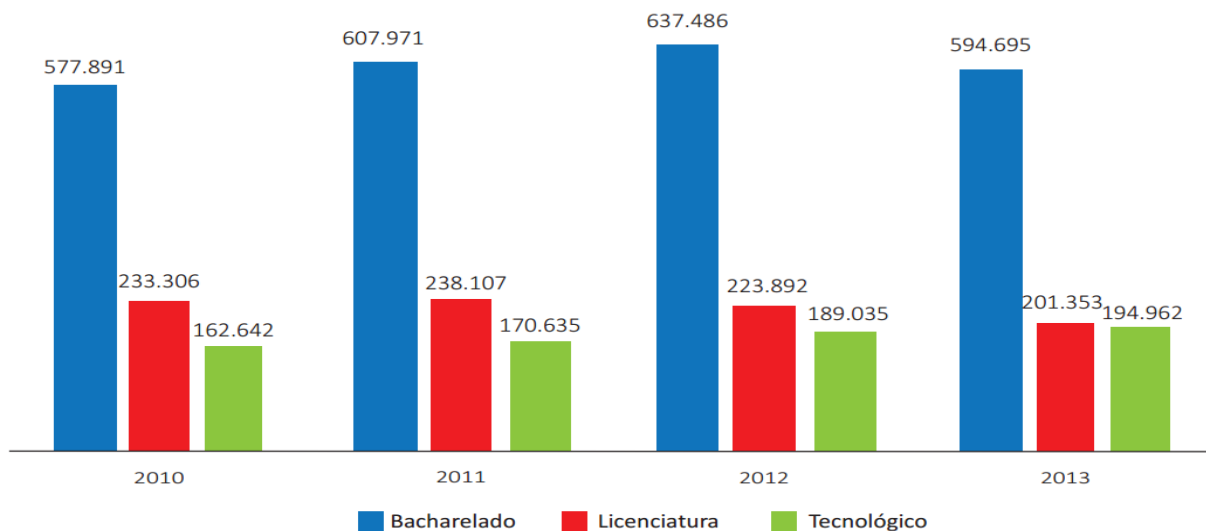


Figura 2.3: Evolução do Número de Concluintes nas IES, por Grau Acadêmico dos Cursos de Graduação. Fonte: INEP [15].

A Tabela 2.5 mostra o percentual de matrículas, ingressantes e concluintes por área de atuação. A área de Ciências Sociais, Negócios e Direito apresenta o maior percentual de matrículas, ingressantes e concluintes. Logo após temos as áreas de Educação e Saúde e Bem Estar Social. As áreas de Engenharia, Produção e Construção; Ciência, Matemática e Computação; Agricultura e Veterinária; possuem um menor percentual de concluintes em relação ao percentual de matrículas e ingressos.

Tabela 2.5: Percentual de Matrículas, Ingressantes e Concluintes por Área de Atuação em 2013. Fonte: INEP [15].

Área de Atuação	Percentual (%)		
	Matrículas	Ingressantes	Concluintes
Ciências Sociais, Negócios e Direito	40,6	41,5	44,3
Educação	18,8	17,2	20,3
Saúde e Bem Estar Social	13,5	12,5	14,1
Engenharia, Produção e Construção	14	14,8	8,2
Ciência, Matemática e Computação	6,1	6,5	5,6
Agricultura e Veterinária	2,4	2,1	1,9
Humanidades e Artes	2,2	2,4	2,7
Serviços	2,3	3,1	2,9

Com base nessas estatísticas, podemos observar que apesar da expansão do acesso ao ensino superior nos últimos anos no Brasil, o número de concluintes nos cursos é inferior ao número de matrículas efetivas. Como exemplo, na área de Ciências, Matemática e Computação, somente 12,5% dos alunos ingressados concluíram o curso, comparado ao total de matrículas efetivas. Segundo Tigrinho [51], apenas o ingresso do aluno nos cursos de graduação não é suficiente para garantir a permanência e a conclusão do aluno no curso.

## 2.3 Evasão em Nível Internacional

Velez [52], em seu estudo realizado para o Centro Nacional de Análises de Dados Longitudinais em Pesquisas Educacionais (National Center for Analysis of Longitudinal Data in Education Research — CALDER), teve como foco uma fração considerável do mercado de educação do Estados Unidos. O estudo avaliou a probabilidade de conclusão de curso para determinados grupos de estudantes.

Segundo o estudo, os administradores e gestores de políticas educacionais devem ter cuidado ao priorizar o ingresso nas faculdades com o objetivo de aumentar a quantidade de alunos formados, visto que ao observar os níveis atuais de preparação acadêmica dos estudantes de graduação, notou-se que muitos não estão preparados para se formarem em um curso de graduação de quatro ou até dois anos.



Para Velez, existem vários motivos para a evasão dos alunos. Por exemplo, a falta de informação sobre o curso ingressado. Os estudantes tomam a decisão de ingressar no ensino superior baseando-se em informações limitadas sobre o curso e sobre as perspectivas futuras, tanto acadêmicas quanto profissionais. Sendo necessário a experiência de pelo menos um ano no curso para obter mais informações. Tal choque, entre o esperado e a realidade, pode causar a evasão precoce do aluno.

Cursos com menor duração (até 2 anos) tiveram índices menores de evasão, em comparação com cursos de maior duração (a partir de 4 anos). Os estudantes que evadiram dos cursos de maior duração tiveram uma probabilidade maior de conclusão de cursos de menor duração, reforçando que é necessário ao aluno um conhecimento prévio sobre o curso.

Foi verificado que apenas 37,9% dos estudantes concluíram o bacharelado em cursos de quatro anos, e que apenas 58,3% conseguiram concluir após seis anos de curso. Logo, acima de 40% dos estudantes de graduação falham em conseguir o diploma de bacharelado em seis anos, e muitos nunca irão completar seus estudos.

O estudo conclui que a maioria dos estudantes que evadiram tiveram desempenho inferior no ensino médio do que os estudantes que se formaram. Indicando que há necessidade de melhor preparação acadêmica anterior, logo no ensino médio.

Além do prejuízo causado na IES devido aos altos índices de evasão, os estudantes perdem anos de experiência e trabalho profissional, podendo também terminar a graduação com grandes dívidas de financiamento escolar. Assim, o investimento pessoal na educação superior também é prejudicado, visto que não há garantia de emprego após a formatura.

Paura e Arhipova [41] estudaram o índice de evasão na Universidade de Agricultura da Latvia (Latvia University of Agriculture — LUA), na Letônia, e concluíram que 35,9% dos alunos evadiram do curso após o primeiro ano.

Foi verificado que o número de estudantes foram diminuindo com o tempo, especificamente após o período de avaliações, visto que os alunos que evadiram tiveram um desempenho ruim e por isso ficaram desmotivados a continuar no curso.

Chies et al. [11] tiveram uma conclusão semelhante na Universidade de Trieste, na Itália. O estudo diz que a maior parte da evasão ocorre no final do primeiro ano após o ingresso no curso, indicando uma provável escolha errada do curso pelo estudante.

Confirmam ainda que os alunos com bom desempenho acadêmico no ensino médio são menos suscetíveis a evadir.

## 2.4 Evasão nas Licenciaturas

Alguns estudos realizados no Brasil, com o objetivo de descobrir informações que expliquem as crescentes taxas de evasão nas Licenciaturas, buscaram investigar as variáveis que estão relacionadas à evasão.

Prietch [45] estudou as causas de evasão nos cursos de Licenciatura da Universidade Federal do Mato Grosso (UFMT). No campus universitário de Rondonópolis, no período de 2001 a 2009, foi realizada uma análise que teve como objetivo avaliar o índice de reprovação nas disciplinas que pertencem à matriz curricular dos cursos. O resultado da análise revelou que há uma grande dificuldade dos alunos em relação às disciplinas obrigatórias de introdução ao curso, que envolvem as matérias de Matemática e Programação. Ainda, os seguintes fatores foram encontrados:

- As disciplinas que possuem o maior índice de reprovação são do primeiro ano: Cálculo 1, Álgebra, Programação e Lógica Matemática, com média de 50,51% de reprovação;
- Há uma menor quantidade de alunos após o segundo ano de curso;
- As taxas de matrículas em disciplinas são maiores no primeiro ano e menores nos anos subsequentes;
- As disciplinas obrigatórias de Projeto e Análise de Algoritmos e de Cálculo Numérico possuem a maior taxa de reprovação entre as disciplinas do terceiro ano;
- A disciplina de Projeto Final e Trabalho de Conclusão de Curso possuem a maior taxa de reprovação no último semestre de curso;
- A explicação sobre a reprovação em disciplinas na área de Exatas foram apontadas como dificuldade de interpretação, métodos de ensino do docente, divergências entre a teoria e prática e avaliação.

No estudo de Andreoli [2], foi elaborado um relatório dos alunos evadidos em todos cursos dos campus da Universidade Federal do Pampa (UNIPAMPA). Os dados obtidos pela Plataforma de Integração de Dados das Instituições Federais de Ensino Superior (PingIFES), em 2010, sobre o campus Bagé, mostrou que o curso de Licenciatura em Matemática obteve o maior número de evadidos. A forma principal de evasão no curso foi por abandono, seguido por trancamento. No relatório, os cursos de graduação do campus que apareceram com mais casos de trancamento foram: Engenharia de Produção, com 33 trancamentos; seguido de Licenciatura em Matemática, com 27; e Engenharia de Computação, com 25. Aqueles que apresentaram mais casos de abandono foram:

Licenciatura em Matemática, com 42 abandonos; seguido de Engenharia de Computação, com 30; Engenharia de Produção, com 22; e Engenharia de Alimentos, com 19. Segundo um questionário aplicado aos alunos, o maior problema relatado foi a dificuldade em disciplinas de Cálculo e o ensino ineficiente do Ensino Médio.

Casartelli [7], em seu estudo envolvendo os alunos evadidos na Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), mostrou que as principais causas da evasão dos alunos na Licenciatura estão relacionadas com a insatisfação sobre o curso, sobre a incompetência dos professores e o relacionamento com os alunos, além de questões financeiras.

Em um estudo sobre a evasão nos cursos de Licenciatura da Universidade Federal de Minas Gerais (UFMG), Lima e Machado [33], por meio de análises bibliográficas, análises de documentos e questionários a coordenadores dos cursos, concluíram que há uma alta tendência de evasão permanente em todos cursos de Licenciatura, pois a atividade profissional do magistério não tem se mostrado atraente no Brasil. Os baixos salários, a falta de segurança nas instituições de ensino, as condições de trabalho precárias, o desprestígio social da atividade e a falta de perspectivas da carreira magisterial contribuem com a desmotivação pela profissão de professor.

Alguns estudos realizados sobre o tema de evasão nos cursos de Licenciatura no Brasil são listados na Tabela 2.6.

Tabela 2.6: Estudos realizados sobre a evasão nos cursos de Licenciatura no Brasil.

<b>Autor</b>	<b>IES</b>	<b>Tema</b>
Gessinger et al. [25]	PUCRS	Estudo com o objetivo de levantar fatores relacionados à evasão
Lima e Machado [33]	UFMG	Pesquisa junto aos coordenadores dos cursos de Licenciatura da universidade com o objetivo de encontrar as causas das evasões
Casartelli [7]	PUCRS	Pesquisa com alunos evadidos e investigação dos motivos da evasão
Fialho e Prestes [21]	UFPB	Investigação da responsabilidade da evasão junto aos gestores educacionais da universidade
Santos [47]	PUCRS	Apresenta dados do Censo da educação e programas do governo com o objetivo de qualificar a mão de obra do país

## 2.5 Evasão na Universidade de Brasília

O Decanato de Ensino de Graduação da UnB (DEG/UnB)<sup>3</sup> é um órgão interno que supervisiona e coordena o ensino nos cursos da UnB. É responsável por desenvolver políticas estudantis de matrícula em disciplina, de avaliação de professor e de estágio para incentivar e garantir a qualidade do ensino.

Segundo o DEG/UnB, na Universidade de Brasília (UnB), além da formatura, a saída de um curso é registrada de várias formas. Existem as seguintes formas de desligamento registradas na SAA: anulação de registro; decisão judicial; falta de documentação; jubileamento; abandono voluntário de curso; não cumprimento de condição; falecimento; transferência; novo vestibular; mudança de habilitação; mudança de curso; mudança de turno; reprovação e/ou abandono pela terceira vez em disciplina obrigatória e abandono.

Uma análise realizada pelo DEG/UnB em 2013 [14], quando o sistema de cotas na UnB completou 10 anos, revelou dados sobre a evasão. No período de 2º/2004 a 2º/2014, 36.896 alunos ingressaram na UnB, 8.472 concluíram o curso, enquanto 10.451 evadiram. Considerando apenas os alunos que saíram da UnB, a porcentagem de formados é de 45%, enquanto de evadidos, 55%. Conforme a Tabela 2.7, que mostra o percentual de alunos desligados nos cursos de Ciências Exatas da UnB no período de 2º/2004 a 2º/2012, é possível observar a discrepância do percentual de formados em relação ao percentual de evadidos.

Tabela 2.7: Evasão nos cursos de Ciências Exatas da Universidade de Brasília. Período de 2º/2004 a 2º/2012.

Curso	Sistema	Matrículas	Desligados	Formados	Cursando
Computação	Cotas	90	34 (38%)	17 (19%)	39 (43%)
Computação	Universal	424	171 (40%)	67 (16%)	186 (44%)
Estatística	Cotas	62	20 (32%)	13 (21%)	29 (47%)
Estatística	Universal	305	68 (22%)	85 (28%)	152 (50%)
Física	Cotas	135	69 (51%)	8 (6%)	58 (43%)
Física	Universal	677	292 (43%)	102 (15%)	283 (42%)
Geologia	Cotas	72	13 (18%)	19 (26%)	40 (56%)
Geologia	Universal	299	46 (15%)	79 (26%)	174 (59%)
Matemática	Cotas	151	72 (48%)	33 (22%)	46 (30%)
Matemática	Universal	740	341 (46%)	135 (18%)	264 (36%)
Química	Cotas	136	36 (26%)	44 (32%)	56 (42%)
Química	Universal	686	205 (30%)	160 (23%)	321 (47%)

Nos cursos de Matemática, considerando apenas os alunos que já saíram do curso, podemos observar que a quantidade de alunos evadidos é superior ao dobro de alunos

<sup>3</sup>[www.unb.br/administracao/decanatos/deg](http://www.unb.br/administracao/decanatos/deg)

formados. No sistema de cotas, de 105 alunos que já saíram do curso, 72 evadiram (69%), enquanto 33 se formaram (31%). Já no sistema universal, de 476 alunos que já saíram do curso, 341 evadiram (72%), enquanto apenas 135 se formaram (28%).

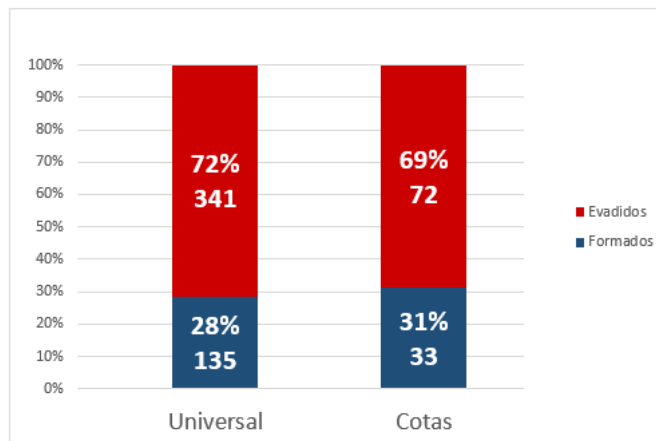


Figura 2.4: A saída no curso de Matemática por sistema de entrada no período de 2º/2004 a 2º/2012. Fonte: DEG/UnB [14].

Na Figura 2.4 é possível observar que não há muita discrepância entre os sistemas de entrada, e que a quantidade de alunos evadidos é superior ao dobro de alunos formados no período de 2º/2004 a 2º/2012.

Bonfim [5], em seu estudo sobre a evasão na UnB, que associou o problema de evasão com as variáveis sociodemográficas, concluiu que o fator idade do aluno é relevante para definir a conclusão do curso, os alunos mais velhos possuem uma chance maior de evadir do que os alunos mais novos. Analisando a trajetória pré-universitária, o fato do aluno já ter concluído um curso superior; ou ter exercido atividade remunerada no período do vestibular; ou ter feito supletivo; ou trabalhar em tempo integral; também aumentam a chance de evasão do aluno.

Em um outro trabalho sobre a evasão na UnB, Lima e Costa [32], utilizaram a técnica de Regressão Logística para encontrar informações precisas sobre a evasão na UnB. Observaram que a variável Sexo foi considerada uma variável importante no modelo final, pois foi constatado que a chance de uma estudante do sexo feminino evadir é quase a metade da chance de um estudante do sexo masculino evadir. A variável Turno não foi considerada significativa no estudo. As idades de ingresso dos estudantes também não foram consideradas significativas para a evasão, mesmo que os estudantes de Licenciatura fossem um pouco mais velhos que os estudantes de Bacharelado. O estudo ainda mostrou que a realização ou não de uma atividade acadêmica extra pode influenciar na evasão. Os estudantes que não foram monitores de disciplinas pelo menos uma vez tiveram três vezes mais chance de evadir do que os estudantes que foram pelo menos uma vez.

Cardoso [6] afirma que a evasão na UnB está relacionada com o rendimento no curso. O baixo rendimento dos alunos está mais ligado à falta de motivação do aluno com o curso do que à dificuldade de acompanhar as disciplinas do curso, em termos de capacidade e domínio dos conteúdos envolvidos. Também constatou em seu estudo que o aluno que trabalha possui o dobro de chance de evadir, conclusão também relatada por Lima e Costa [32].

Outro estudo realizado por Florencio [23], sobre a evasão no curso de Licenciatura em Computação da UnB, concluiu que algumas disciplinas obrigatórias estão associadas com o problema de evasão no curso. Das disciplinas obrigatórias observadas, as que mais influenciaram no fator de evasão do aluno pertencem ao primeiro e segundo semestre do curso: Lógica Computacional, Cálculo 1, Computação Básica e Estrutura de Dados.

# Capítulo 3

## Mineração de Dados

Esse capítulo é destinado à apresentação de conceitos gerais sobre mineração de dados. A Seção 3.1 introduz conceitos gerais sobre dado, informação e conhecimento. A Seção 3.2 apresenta as principais definições de mineração de dados. A Seção 3.3 apresenta os principais processos de extração de conhecimento pela mineração de dados. A Seção 3.4 define técnicas de mineração de dados. Por fim, a Seção 3.5 apresenta a ferramenta usada neste trabalho, denominada WEKA.

### 3.1 Dado, Informação e Conhecimento

Gordon [26] define que dados são quaisquer fatos, números ou textos que podem ser processados pelo computador. Metadados são dados sobre os dados, ou seja, são a descrição dos dados. Informações são todos os padrões, associações ou relações que os dados podem fornecer. Conhecimento são informações sobre padrões históricos e tendências futuras.

Angeloni [3] explica que os dados, a informação e o conhecimento são elementos fundamentais para a tomada de decisão nas organizações e constituem um sistema hierárquico de difícil delimitação. Dados podem ser vistos de diferentes perspectivas, a depender do indivíduo que os interpreta. Para um indivíduo pode ser informação, para outro pode ser conhecimento. Os dados são considerados a matéria-prima da informação. São elementos brutos desvinculados de significado. Já a informação consiste em dados processados, contextualizados, providos de significado, relevância e propósito. O conhecimento pode ser considerado como o conjunto de informações mais valiosas que são processadas pelo indivíduo.

Davenport [13] enfatiza que a definição dos termos é complexa e afirma que a distinção entre eles é imprecisa. No máximo, pode-se elaborar um processo. Dados são simples observações sobre o estado do mundo, fatos brutos ou entidades quantificáveis. O conceito informação é um conjunto de dados dotados de relevância e propósito, que são analisados

pelo ser humano. O conhecimento consiste em informações valiosas da mente humana, que inclui reflexão, síntese e contexto. É definido como o conjunto de informações que foram acrescentadas de interpretação. A Tabela 3.1 apresenta as características dos termos.

Tabela 3.1: Dados, informação e conhecimento, segundo Davenport [13].

Dados	Informação	Conhecimento
Facilmente estruturados	Requer unidade de análise	Difícil estruturação
Facilmente obtidos	Exige consenso em relação ao significado	Difícil captura em máquinas
Facilmente transferíveis	Exige interpretação humana	Difícil transferência

## Os tipos de dados

Os conjuntos de dados podem diferir de diversas maneiras e conter características especiais. Os atributos utilizados para descrever os dados podem ser quantitativos ou qualitativos. Os tipos de dados presentes nos conjuntos de dados determinam quais ferramentas e técnicas podem ser utilizadas para analisar os dados.

## Qualidade dos Conjuntos de Dados

O entendimento e o aperfeiçoamento da qualidade dos dados pode aumentar significativamente a qualidade do resultado final da mineração e análise dos dados. A presença de ruídos, valores atípicos, omissões, inconsistências ou duplicações são fatores comumente encontrados em conjunto de dados gerais, que podem comprometer o fenômeno ou características da população em estudo. Minimizar esses problemas pode aumentar a qualidade geral do resultado.

## 3.2 Definições de Mineração de Dados

Kumar [40] explica que a mineração de dados é uma tecnologia que acopla métodos tradicionais de análises de dados com algoritmos sofisticados para o processamento de grandes volumes de dados.

Pela definição de Gordon [26], a mineração de dados é o processo de analisar dados sobre diferentes perspectivas e resumir em informações úteis. Tecnicamente, é o processo de encontrar correlações ou padrões entre vários campos presentes em grandes bancos de dados relacionais.



Witten et al. [54] explica que a mineração de dados é definida como o processo de descoberta de padrões em dados, que devem ser significativos para atingir um entendimento sobre eles. O grande volume de dados armazenados aumenta a dificuldade e a complexidade de entendimento sobre eles e consequentemente também aumentam as chances de descobertas de padrões não evidentes.

Han e Kamber [27] definem a mineração de dados como o resultado da evolução natural da tecnologia da informação e afirma que enormes quantidades de dados estão sendo acumuladas devido à expansão tecnológica da informação, que se iniciou por dois fatores principais: o rápido crescimento da capacidade computacional de armazenamento e gerenciamento de dados e a computadorização da sociedade em muitos setores.

Na visão de Fayyad et al. [20], o processo de descoberta de conhecimento em banco de dados refere-se a todo processo de descoberta de conhecimento útil nos dados, enquanto a mineração de dados refere-se a uma etapa particular nesse processo. A mineração de dados é a aplicação de algoritmos específicos para extrair padrões de dados.

### 3.3 Processos de Extração de Conhecimento em Dados

Segundo Fayyad et al. [20], o método tradicional de transformar dados em conhecimento baseia-se na análise e interpretação manual, sendo ele demorado, de alto custo e altamente subjetivo. A depender da complexidade e da quantidade em que os dados se apresentam, torna-se inviável ou muitas vezes impossível para um ser humano utilizar o método tradicional.

Fayyad et al. [20] determinam que a mineração de dados envolve encaixar modelos para determinar padrões nos dados observados e propõe o processo de descoberta de conhecimento em banco de dados (*Knowledge Discovery in Databases/KDD*).

O KDD é um processo de identificação de novos padrões, potencialmente úteis e altamente compreensíveis, que inclui adicionalmente atividades multidisciplinares (mostrado na Figura 3.1) como a seleção, o pré-processamento, a transformação, a mineração de dados e a análise interpretativa dos resultados obtidos. Envolve cinco passos:

1. Seleção: criação de um conjunto de dados relevantes e determinação um subconjunto de variáveis ou amostra de dados;
2. Pré-processamento: limpeza dos dados, filtragem, remoção de ruídos e valores extremos. Estratégias devem ser determinadas para não comprometerem a qualidade da análise;

3. Transformação: transformação dos dados brutos usando a redução dimensional para encontrar representações invariáveis dos dados;
4. Mineração: busca de padrões de interesse a depender do objetivo. Transformação dos dados em informações significativas;
5. Análise Interpretativa: análise e interpretação dos dados, avaliação dos resultados obtidos pela mineração de dados.

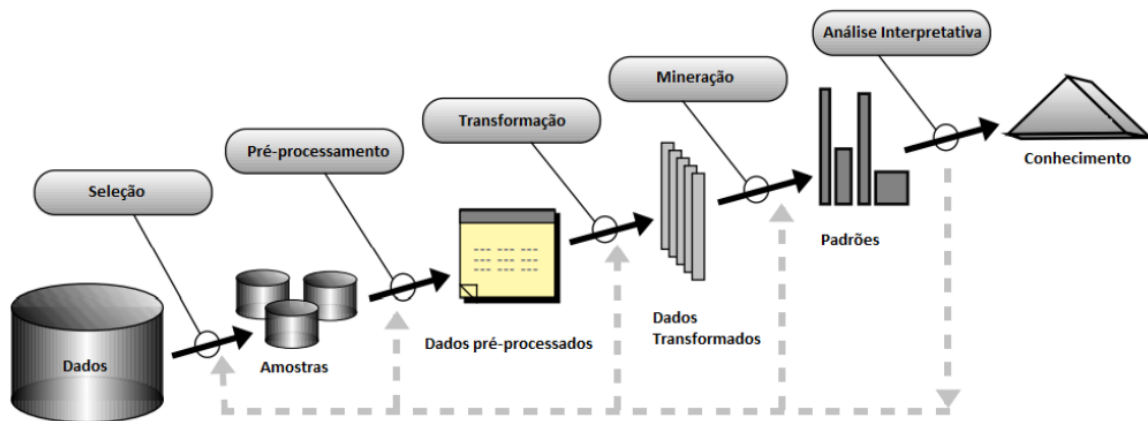


Figura 3.1: Uma visão geral das etapas que compõem o processo de descoberta de conhecimento em banco de dados (*Knowledge Discovery in Databases/KDD*) (Adaptado de Fayyad et al. [20]).

Olson e Delen [39] enfatizam dois processos de mineração de dados, o Processo Padrão Inter-Indústrias para Mineração de Dados (*Cross Industry Standard Process for Data Mining - CRISP*) e o processo Experimental, explorar, modificar, modelar, avaliar (*Sample, explore, modify, model, assess - SEMMA*).

Chapman et al. [9] definem o processo de mineração de dados CRISP em seis fases, conforme mostrado na Figura 3.2. Em detalhes, as seis fases do processo CRISP são:

1. Entendimento dos negócios: concentra-se em entender os objetivos e requisitos do projeto a partir de uma perspectiva de negócios e, em seguida, converter esse conhecimento em uma mineração de dados definindo os problemas e os objetivos a serem alcançados;
2. Entendimento dos dados: consiste em uma coleta inicial de dados e atividades para familiarização com os dados para identificar problemas de qualidade;

3. Preparação dos dados: consiste em atividades para construção do conjunto final dos dados a partir dos dados brutos. As atividades incluem seleção de atributos, filtragem, transformação e limpeza de dados;
4. Modelagem: várias técnicas de modelagem são selecionadas e aplicadas. Tipicamente, existem diversas técnicas para o mesmo problema de mineração de dados. Algumas técnicas dependem da forma dos dados, portanto voltar à etapa anterior para a preparação dos dados é muitas vezes necessário;
5. Avaliação: após concluir a fase de modelagem é necessário analisar, validar e testar os resultados obtidos para avaliar o desempenho e a confiabilidade do modelo selecionado. É preciso verificar se os resultados cumprem o objetivo estabelecido;
6. Distribuição e entrega: esta fase consiste na geração de um relatório de boa qualidade, visualmente prático e acessível a todos envolvidos no processo.

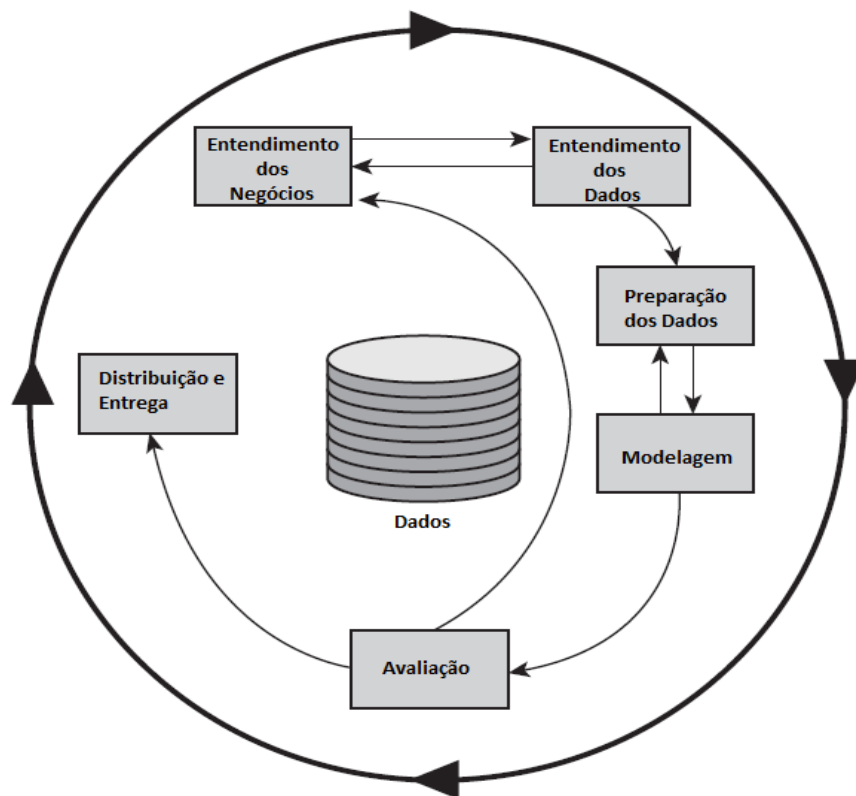


Figura 3.2: Uma visão geral das etapas que compõem o processo CRISP (Adaptado de Chapman et al. [9]).

Olson e Delen [39] definem o processo de mineração SEMMA em cinco etapas, conforme mostrado na Figura 3.3. Detalhadamente as seguintes etapas fazem parte do processo SEMMA:

1. Amostragem: consiste em criar uma amostra de dados que contém informações significativas;
2. Exploração: deve-se explorar os dados buscando relacionamentos antecipados e anomalias dos dados, para obter maior entendimento e idéias de exploração;
3. Modificação: consiste em modificar os dados criando, selecionando e transformando as variáveis focando o processo da seleção de um modelo;
4. Modelagem: consiste em modelar os dados usando ferramentas analíticas para buscar combinações de dados que cumprem o objetivo estabelecido;
5. Avaliação: avaliação dos resultados obtidos considerando a utilidade e consistência dos dados.

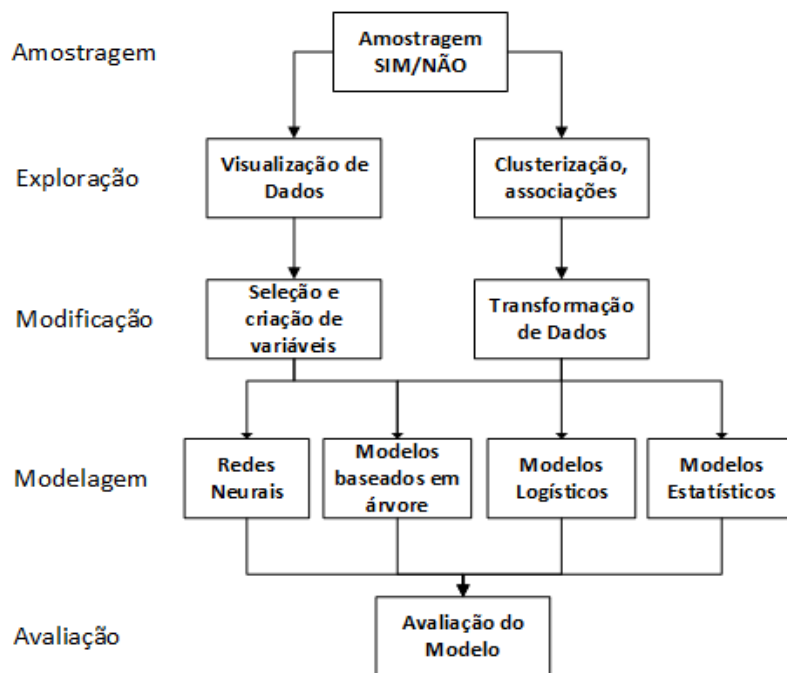


Figura 3.3: Uma visão geral das etapas que compõem o processo SEMMA. (Adaptado de Olson e Delen [39]).

### 3.4 Técnica de Mineração de Dados

Existem diversas técnicas de mineração de dados que podem ser utilizadas nos processos de mineração de dados. Cada técnica possui um modo de disponibilizar, tratar e organizar os dados. Cabe ao usuário avaliar quais técnicas se adequam às necessidades e ao contexto da mineração de dados.

## Classificação

A técnica de classificação, explicada por Donalek [18], é a tarefa de atribuir dados para uma ou várias categorias pré-definidas. A técnica divide o conjunto de dados de treinamento em classes pré-definidas e então utiliza um subconjunto de dados de teste para testar o modelo de classificação encontrado. Após a realização do teste, é possível detectar regras e padrões no conjunto de dados, e assim prever a qual classe o dado pertence.

Han e Kamber [27] definem a técnica de classificação como um processo que é realizado em duas fases. A primeira fase, denominada treinamento, consiste na criação de um conjunto de regras de classificação a partir dos dados de treinamento. Nessa fase, para cada atributo presente no conjunto de dados de treinamento, é gerado uma regra que determinará a classe de cada instância de dados de acordo com o valor de cada atributo. O conjunto de regras de classificação compõe o modelo de classificação. A Figura 3.4 ilustra um exemplo da primeira fase da classificação. Nesse exemplo, para cada atributo presente no conjunto de dados de treinamento, o classificador cria uma regra de classificação para determinar se um empréstimo é seguro ou de risco.

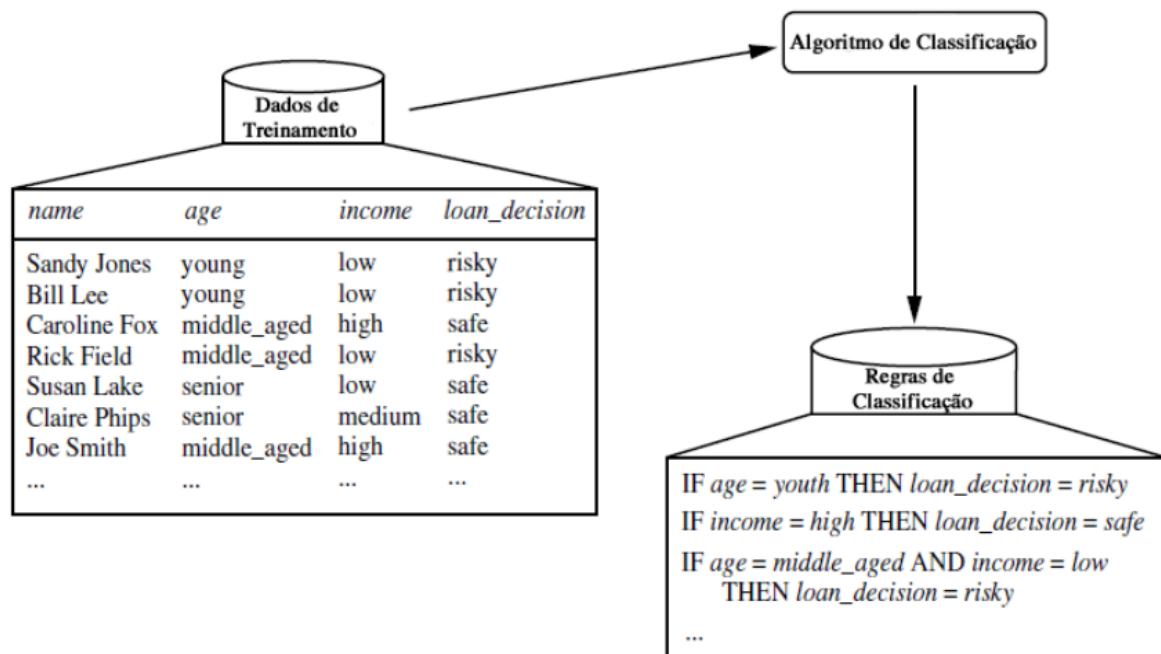


Figura 3.4: Primeira etapa de classificação de acordo com Han e Kamber [27].

A segunda fase, denominada teste, é quando ocorre efetivamente a classificação dos dados a partir do conjunto de dados de teste. Nessa fase é importante estimar a acurácia do modelo de classificação, ao observar a porcentagem de classificações corretas e incorretas. Caso a quantidade de classificações corretas sejam suficientes para a aplicação, ou seja,

caso o modelo de classificação tenha boa acurácia, é possível utilizá-lo para predição de classes do conjunto de dados estudado.

A Figura 3.5 ilustra um exemplo da segunda fase da classificação. No exemplo, o classificador utiliza o conjunto de teste para executar um conjunto regras, que estão presentes no modelo de classificação, para predizer se um empréstimo é seguro ou de risco.

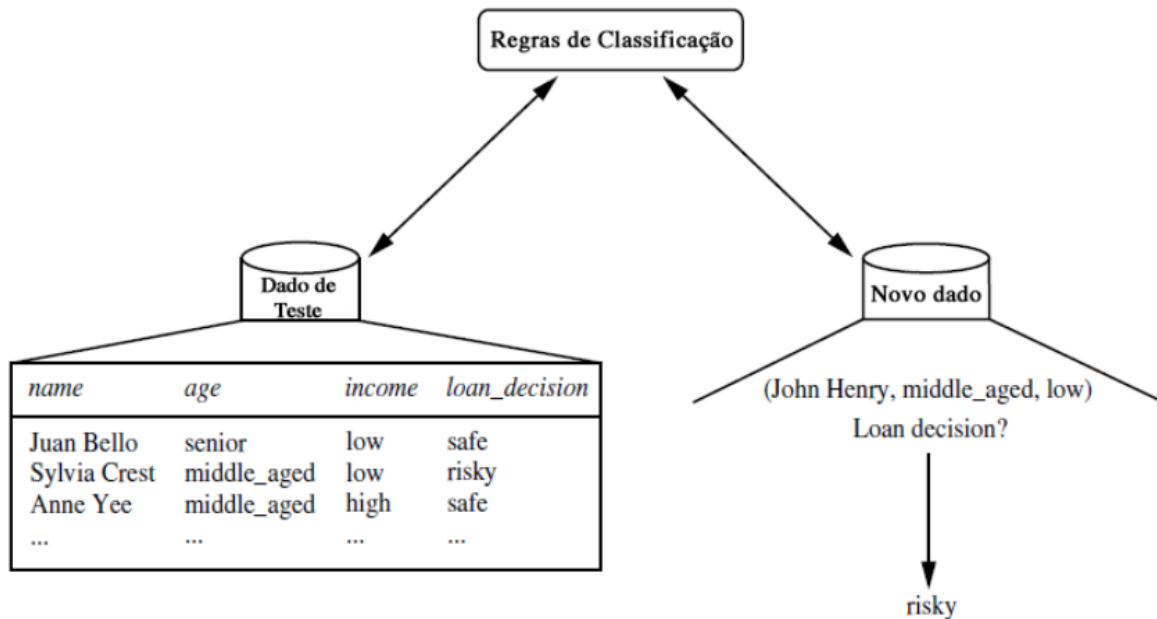


Figura 3.5: Segunda etapa de classificação de acordo com Han e Kamber [27].

Para esse trabalho, será utilizada a técnica de **classificação**, pois como visto, a classificação pode ser utilizada para a predição de classes de um conjunto estudado. Em específico, utilizaremos a técnica de classificação no conjunto de dados dos alunos para verificar quais as disciplinas que estão ligadas à evasão no curso de Licenciatura em Matemática (Diurno) da UnB.

## Algoritmo de Classificação

Para realizar a técnica de classificação, é preciso escolher um algoritmo classificador. Entre os melhores algoritmos de classificação apresentados por Wu et al. [56], está o algoritmo **C4.5**, desenvolvido por Quinlan [46].

Segundo Quinlan [46], o algoritmo tem como objetivo gerar um modelo classificador na forma de uma árvore de decisão, utilizando a técnica de busca gulosa, na qual a cada passo determina qual atributo é o mais preditivo e cria uma sub-árvore com base nesse atributo.

Para esse trabalho, será utilizado o algoritmo de classificação **C4.5**.

### 3.5 WEKA

Waikato Environment for Knowledge Analysis (WEKA) é uma coleção de algoritmos de aprendizagem de máquina sofisticados e ferramentas de pré-processamento de dados. Ele fornece suporte extensivo para todo processo experimental de mineração de dados, incluindo a preparação dos dados de entrada, avaliação da aprendizagem de esquemas estatísticos, visualização da entrada de dados e resultado da aprendizagem.

A ferramenta inclui métodos para os problemas principais da mineração de dados: regressão, classificação, clusterização, regras de associações e seleção de atributos. A ferramenta permite a importação de arquivos em diversos formatos, um deles é o formato padrão do WEKA, o ARFF (*Attribute-Relation File Format*), que foi utilizado nesse trabalho. O formato completo do banco de dados utilizado encontra-se no Anexo I.



Figura 3.6: Interface principal WEKA - Escolha de aplicações.

Como aplicações, o WEKA disponibiliza o *Explorer*, o *Experimenter*, o *KnowledgeFlow* e o *Simple CLI*.

O *Explorer*, mostrado na Figura 3.7, é a aplicação principal do WEKA, que fornece fácil acesso às funcionalidades de pré-processamento de dados, classificação, clusterização, associação, seleção de atributos e visualização de resultados.

Em específico, cada aba é responsável por fornecer as seguintes funcionalidades:

- *Preprocess*: escolha e modificação do conjunto de dados a ser trabalhado;
- *Classify*: treinamento e avaliação de esquemas de aprendizagem para realizar classificação ou regressão;
- *Cluster*: realiza a clusterização dos conjuntos de dados;
- *Associate*: aprende e avalia regras de associação pelo conjunto de dados;

- *Select Attributes*: seleciona os atributos mais relevantes do conjunto de dados;
- *Visualize*: permite a visualização cartesiana de dados a partir de atributos selecionados.

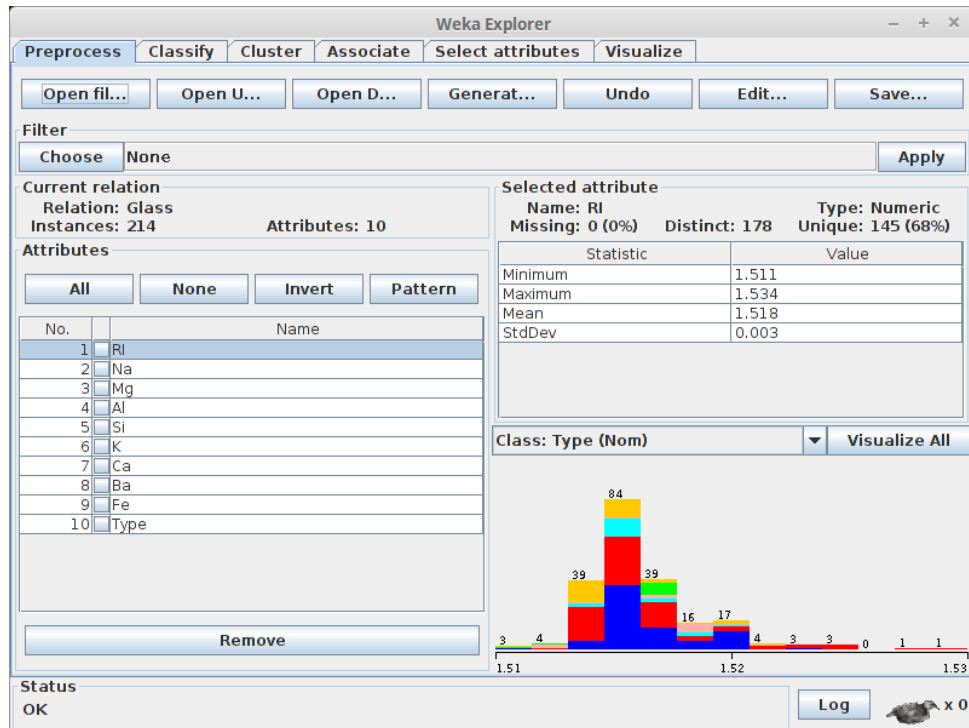


Figura 3.7: Interface *Explorer* do WEKA.



# Capítulo 4

## Informações do Curso e dos Alunos

Neste capítulo, são apresentadas informações do curso de Licenciatura em Matemática (Diurno) e dados dos alunos, que foram usados nas análises. Na Seção 4.1 é apresentado o curso de Licenciatura em Matemática (Diurno) e, em seguida, na Seção 4.2 são apresentadas informações sobre os dados dos alunos.

### 4.1 Licenciatura em Matemática (Diurno)

De acordo com o Portal da UnB [12], o curso de Matemática tem um campo amplo de aplicações, as mais comuns são na Computação, na Física, na Biologia, na Economia e na Matemática Financeira, permitindo assim a especialização em diversas áreas de conhecimento. O curso presencial de Matemática da UnB possui duas habilitações: o bacharelado e a licenciatura plena. O ensino da Matemática, em ambas habilitações, consiste no desenvolvimento de cinco áreas básicas: Álgebra, Análise, Geometria, Probabilidade e de Teoria dos Números. No caso da habilitação em licenciatura plena, o aluno recebe uma formação pedagógica própria, e é capacitado para atuar como professor no ensino básico. Os dados contendo informações sobre o curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília foram obtidos pelo site Matricula Web<sup>1</sup>.

Apresentamos inicialmente informações gerais sobre o curso de Licenciatura em Matemática (Diurno). Em seguida, são apresentadas as disciplinas obrigatórias do curso e por fim, mostramos o fluxo de habilitação vigente do curso.

#### 4.1.1 Descrição do Curso

A Licenciatura em Matemática (Diurno) é uma habilitação do curso de Matemática que permite aos licenciados o exercício do magistério no ensino básico. Na Universidade de

---

<sup>1</sup>[matriculaweb.unb.br](http://matriculaweb.unb.br)

Brasília, o curso possui duas habilitações, divididas em dois turnos: diurno e noturno. Essas se diferenciam em apenas dois parâmetros: A quantidade máxima de créditos por período e o limite mínimo de permanência semestral. A Tabela 4.1 mostra as informações sobre a habilitação em Licenciatura em Matemática (Diurno).

Tabela 4.1: Licenciatura em Matemática (Diurno) - Informações sobre a habilitação de acordo com o Matricula Web [17].

<b>Licenciatura em Matemática Diurno</b>	
Nome do curso:	Matemática
Habilitação:	Licenciatura em Matemática
Grau:	Licenciado
Código da Habilitação:	1325
Unidade Acadêmica:	Instituto de Ciências Exatas (IE)
Campus:	Darcy Ribeiro
Turno:	Diurno
Reconhecido pelo MEC:	Sim
Currículo vigente em:	2014/1
Mínimo de créditos por período:	16
Máximo de créditos por período:	32
Limite mínimo de permanência:	6
Limite máximo de permanência:	14
Créditos exigidos para formatura:	188
Quantidade máxima de créditos no Módulo Livre:	24

### 4.1.2 Disciplinas Obrigatórias

As disciplinas obrigatórias para ambas as habilitações, diurno e noturno, não diferem. Para obter a quantidade necessária de créditos para conclusão do curso e a formatura, o aluno precisa, além de obter a aprovação em todas as disciplinas obrigatórias, complementar a quantidade necessária obtendo aprovação em disciplinas optativas ou no módulo livre.

O currículo atual da habilitação de Licenciatura em Matemática (Diurno) está baseado no Projeto Pedagógico de Curso (PPC) vigente em 1º/2014. Desde a criação do curso de Licenciatura em Matemática (Diurno), em 2º/1977, o PPC foi alterado duas vezes. Em cada alteração, tanto o fluxo quanto o currículo do curso foram alterados. A primeira alteração foi realizada em 1º/2005 (disponível no Anexo II). A segunda e atual foi realizada em 1º/2014. A Tabela 4.2 mostra o currículo vigente a partir de 1º/2014.

Para esse estudo, iremos considerar apenas as **disciplinas obrigatórias do currículo vigente atual**, pois as disciplinas que foram retiradas desse currículo serão irrelevantes para determinar se um aluno estará em risco de evasão.

Tabela 4.2: Currículo da Licenciatura em Matemática - Disciplinas obrigatórias de acordo com o Matricula Web [17].

<b>Código - Disciplina</b>	<b>Quantidade de Créditos</b>
113107 - ALGEBRA 1	4
117421 - ÁLGEBRA PARA O ENSINO 1	6
117501 - ÁLGEBRA PARA O ENSINO 2	6
113204 - ANALISE 1	4
125172 - APRENDIZAGEM NO ENSINO	4
113034 - CALCULO 1	6
113042 - CALCULO 2	6
113051 - CALCULO 3	6
113824 - CALCULO DE PROBABILIDADE 1	6
125156 - DESENVOL PSICOLOGICO E ENSINO	4
194531 - DIDATICA FUNDAMENTAL	4
118001 - FISICA 1	4
118010 - FISICA 1 EXPERIMENTAL	2
118028 - FISICA 2	4
117161 - GEOMETRIA 1	4
117170 - GEOMETRIA 2	4
105881 - GEOMETRIA ANALÍTICA MATEMÁTICA	4
117471 - GEOMETRIA PARA O ENSINO 1	6
117480 - GEOMETRIA PARA O ENSINO 2	6
113093 - INTRODUCAO A ALGEBRA LINEAR	4
113913 - INTRODUCAO A CIEN COMPUTACAO	4
194221 - ORGANIZACAO DA EDUCACAO BRASILEIRA	4
117510 - REGÊNCIA 1	8
117439 - REGÊNCIA 2	8
113115 - TEORIA DOS NUMEROS	4
113069 - VARIABEL COMPLEXA 1	6
<b>Total de créditos obrigatórios</b>	<b>128</b>
<b>Total de disciplinas obrigatórias</b>	<b>26</b>

### 4.1.3 O Fluxo Vigente da Habilitação Diurna

O fluxo da habilitação atual, assim como o currículo, está baseado no PPC vigente em 1º/2014. A Figura 4.1 mostra o fluxo de habilitação atual do curso de Licenciatura em Matemática (Diurno).

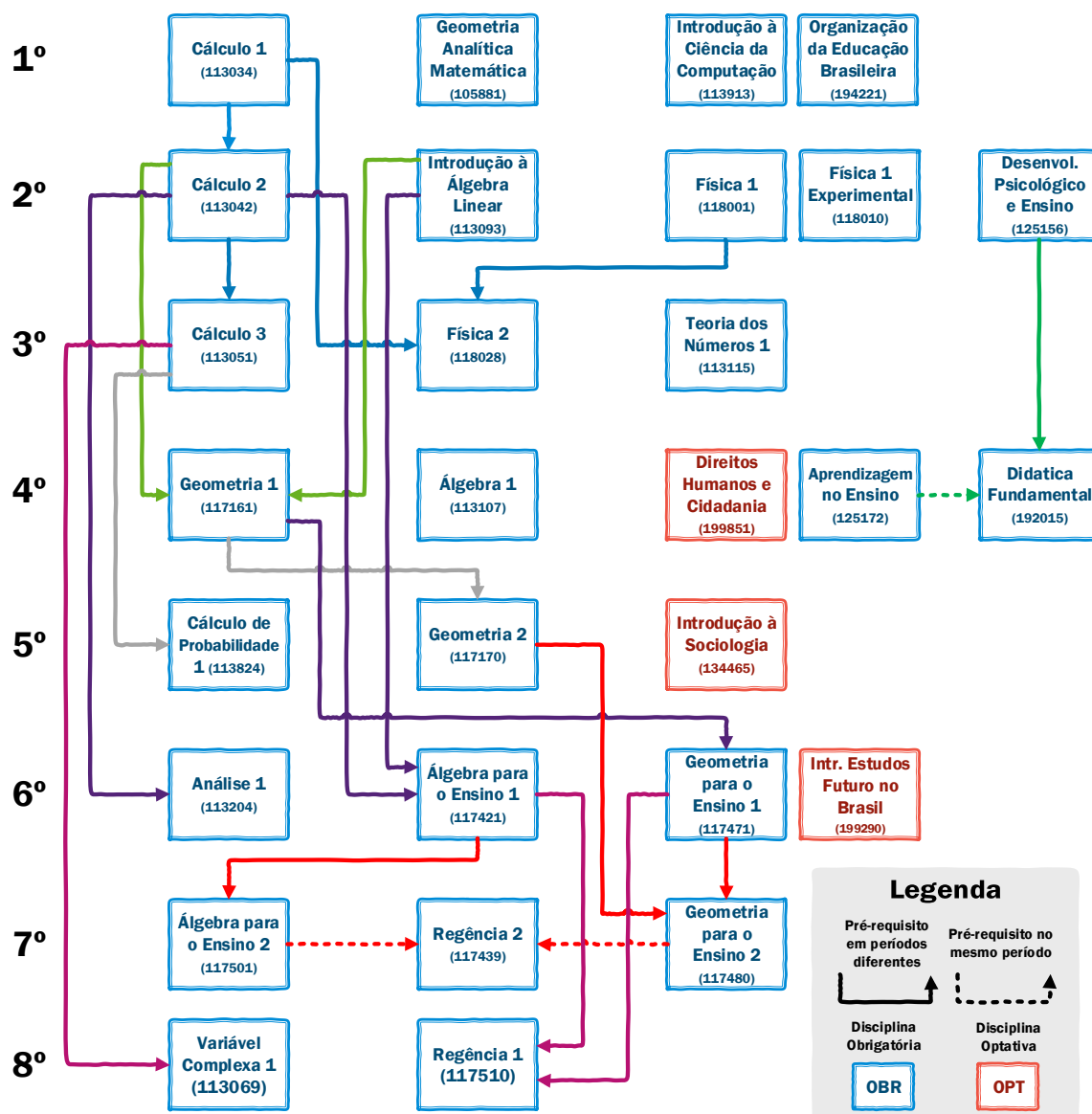


Figura 4.1: Fluxo de disciplinas do curso de Licenciatura em Matemática (Diurno) da Universidade de Brasília [17].

## 4.2 Dados dos Alunos

Os dados sobre os alunos, utilizados como base para a análise estatística neste trabalho, foram cedidos pelo Decanato de Planejamento e Orçamento da UnB (DPO/UnB) através do Sistema de Informação de Graduação (SIGRA), que é mantido pelo Centro de Informática da UnB (CPD/UnB). Neles estão contidos tanto dados pessoais quanto acadêmicos de todos 1.390 alunos ingressados desde a criação do curso em 2º/1977.

Inicialmente apresentamos todos os metadados sobre os dados brutos, e em seguida são mostradas como foram feitas a limpeza e a filtragem dos dados brutos.

### 4.2.1 Dados Brutos

A Tabela 4.3 mostra o total de 76 atributos presentes nos dados brutos em arquivo no formato Valores Separados por Vírgula (*Comma Separated Values – CSV*). É possível obter informações pessoais sobre os alunos, o seu histórico acadêmico, o curso e a habilitação.

Tabela 4.3: Os atributos presentes nos dados brutos

Atributo	Descrição
MatricAluno	A matrícula do aluno no curso
Sexo	O gênero do aluno
Nível	O nível de formação
Nacionalidade	O país de origem do aluno
Nascimento	A data de nascimento
País	O país atual do aluno
UF	A Unidade Federativa do aluno
Cota	O sistema de entrada
PNE	Se é portador de necessidades especiais
TipoEscola	A iniciativa da escola: Particular ou Pública
Raca	Informa a raça do aluno
Local de Nascimento	A cidade e estado de nascimento
Identificador	Atributo desconhecido, apresenta um número atribuído ao aluno
CodOpcao	Informa o código da opção de habilitação do aluno
Prioridade Opcao	A prioridade da opção de escolha da habilitação
Ano Ingresso	O ano de ingresso na instituição
Semestre Ingresso	O semestre de ingresso na instituição
Forma Ingresso	A forma de ingresso: Vestibular, PAS, ENEM...

Ano Saida	Informa o ano de saída da instituição, caso disponível
Semestre Saida	Informa o semestre de saída da instituição, caso disponível
Forma Saida	Informa a forma de saída: Formatura, Desligamento, Está Cursando...
Ano Ingresso Opção	O ano de ingresso na habilitação
Semestre Ingresso Opção	O semestre de ingresso na habilitação
Forma Ingresso Opção	A forma de ingresso na habilitação: Dupla Habilitação, Vestibular...
Ano Data	A data de entrada na instituição, caso disponível
Mes Ingresso Opcao	O mês de ingresso na habilitação
Ano Saida Opcao	O ano de saída da habilitação
Semestre Saida Opcao	O semestre de saída da habilitação
Forma Saida Opcao	A forma de saída da habilitação
Aluno Registrado	Caso aluno foi registrado na instituição
Periodo Curricular	O semestre e o ano referente ao período curricular
Creditos	A quantidade de créditos cursados pelo aluno
NumCreditos	A quantidade de créditos necessários para formatura na habilitação
Judicial	Caso o aluno tenha ingressado a partir de um processo judicial
Polo Aluno	Polo de curso à distância
IRA	O Índice de Rendimento Acadêmico
Maior Periodo Formando	O maior período em que o aluno estava formando
Nome Opcao	Nome da Opção (Habilitação)
Ano Inicio Opcao	O ano de vigência do PPC da habilitação
Semestre Inicio Opcao	O semestre de vigência do PPC da habilitação
Ano Fim Opcao	O ano do fim da vigência do PPC da habilitação
Semestre Fim Opcao	O semestre do fim da vigência do PPC da habilitação
Creditos Formatura	Quantidade de créditos necessários para formatura
CodCurso	Código do curso: 141
Nivel Opcao	O nível da opção: Graduação
Tempo Minimo	O tempo mínimo em semestres para conclusão do curso
Tempo Maximo	O tempo máximo em semestres para conclusão do curso
Duracao	O tipo de duração
Data Resolucao	A data de resolução do curso

Horas Em Extensao	Quantidade de horas em extensão necessárias para conclusão
Ano Inicio Curso	O ano de início do curso
Semestre Inicio Curso	O semestre de início do curso
Ano Fim Curso	O ano de fim do curso
Semestre Fim Curso	O semestre de fim do curso
Nome Curso	O nome do curso
Cod Inst	O código institucional
Cod Depar	O código do departamento
Area de Conhecimento	A área de conhecimento
Nivel Curso	O nível do curso
Duracao Curso	A duração do curso
Turno Curso	O turno do curso
Forma Curso	A forma do curso
ENADE	O número ENADE do curso
Cód. ENADE	O código ENADE do curso
Grau	O gráu do curso: Graduação
Modalidade	A modalidade do curso: Educação Presencial
NomeL	O nome simples do curso: MATEMÁTICA
Duração Padrão Semestres	A duração padrão do curso em semestres: 8
Turno	O turno do curso
LocalOferta	O campus do curso: Darcy
Nome Anuário	O nome do curso
Unidade-Cód	O código da unidade do curso
Unidade	A unidade do curso: IE
CodDisc[ano][semestre]	O código da disciplina em questão por ano e semestre
Mencao[ano][semestre]	A menção do aluno na disciplina em questão por ano e semestre
Nome Disciplina [ano][semestre]	O nome da disciplina em questão por ano e semestre
<b>Total</b>	<b>76 atributos</b>

#### 4.2.2 Filtragem e Limpeza dos Dados Brutos

Ao realizar uma análise mais concisa sobre os dados brutos, foi necessário filtrar os dados, removendo atributos que não eram significativos para o estudo, assim como atributos que

continham valores redundantes.

Os seguintes atributos que sempre possuíam o mesmo valor foram removidos: Unidade-Cód: 11, Unidade: IE, LocalOferta: DARCY, Duração Padrão Semestres: 8, NomeL: MATEMATICA, Modalidade: Educacao Presencial, Grau: Licenciatura, Forma Curso: 0, Area de Conhecimento: 27, Cod Depar: Departamento de Matematica, Cod Inst: Instituto de Ciencias Exatas, Nome Curso: Matematica, Semestre Fim curso: 9, Ano fim curso: 9999, semestre inicio curso: 1, Horas em extensao: 0, Data Resolucao: 13/03/1973, DURACAO: 'Duracao Plena DP' Cod Grau: 'Licenciado', Tempo Maximo: '14', Nivel Opcao: 'Graduacao', Creditos Formatura: '188', Semestre fim Opcao: '9', Ano fim opcao: '9999', Semestre inicio opcao: '1', Ano inicio opcao: '2014', Nome opcao: 'LICENCIATURA EM MATEMATICA', Aluno Registrado: 'Sim'.

Após a remoção, foi realizada a limpeza de dados, removendo caracteres especiais e normalizando a codificação em ASCII (*American Standard Code for Information Interchange*). O algoritmo utilizado para a limpeza dos dados está no Anexo III.

Logo em seguida, todos os dados foram inseridos em um banco de dados MySQL<sup>2</sup> para facilitar na consulta, na correlação e no entendimento dos dados, assim como também fornecer suporte no processo de mineração de dados. O algoritmo utilizado para a inserção dos dados no banco de dados MySQL, assim como as consultas, está disponível no Anexo IV.

Por fim, os atributos selecionados para a mineração de dados foram: MATRICALUNO, SEXO, NASCIMENTO, UF, TIPO\_ESCOLA, RACA, ANO\_INGRESSO\_OPCAO, SEMESTRE\_INGRESSO\_OPCAO, FORMA\_INGRESSO\_OPCAO, FOR\_SAIDA\_OPCAO, ANO\_SAIDA\_OPCAO, SEMESTRE\_SAIDA\_OPCAO, IRA.

Além desses atributos, um novo atributo chamado SAIDA foi criado com base no atributo FOR\_SAIDA\_OPCAO. Esse atributo foi utilizado para classificar, de forma mais eficiente, a saída do curso pelo aluno. Caso o valor do atributo FOR\_SAIDA\_OPCAO seja relacionado à evasão, então o valor do atributo SAIDA será 'Evadiu'. Caso contrário será 'Formou'.

---

<sup>2</sup>[www.mysql.com](http://www.mysql.com)



# Capítulo 5

## Análises Estatísticas

As análises estatísticas presentes neste capítulo foram realizadas após a etapa de filtragem e limpeza dos dados brutos. O objetivo dessas análises é fornecer informações relevantes para identificar os motivos que podem estar relacionados com a evasão do aluno. A Seção 5.1 apresenta a situação atual dos alunos. A Seção 5.2 mostra as diferentes formas e estatísticas de ingresso no curso. Já a Seção 5.3 apresenta as principais formas de saída e estatísticas de evasão do curso. Por fim, a Seção 5.4 apresenta as taxas de reprovação das disciplinas obrigatórias que apresentam as maiores taxas de reprovação.

### 5.1 A Situação Atual

Atualmente, de 1.390 alunos, 159 (11%) estão cursando, 460 (33%) se formaram e 771 (55%) evadiram. A Figura 5.1 mostra o gráfico da situação atual.

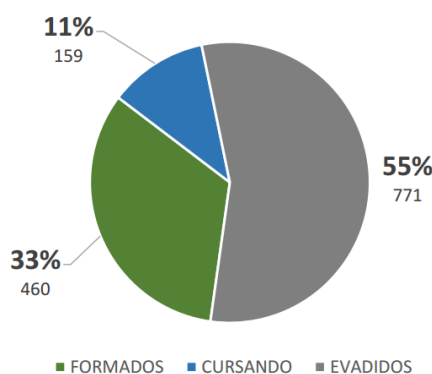


Figura 5.1: Situação atual dos alunos analisados (2º/1977 a 1º/2016).

## 5.2 Formas de Ingresso no Curso

Ao observar a Figura 5.2, que mostra o gráfico da quantidade de alunos ingressados por forma de ingresso no período de 2º/1977 a 1º/2016, podemos observar que a maior quantidade de alunos ingressou no curso através dos meios tradicionais como o Vestibular (887 – 64%) e pelo PAS (176 – 13%). Em seguida, o ingresso foi feito pela opção de Dupla Habilitação (104 – 7%), em que o aluno escolhe fazer o curso de Licenciatura, após cursar o Bacharelado em Matemática. A partir do primeiro semestre de 2014, a UnB adotou, como método de ingresso, o Sistema de Seleção Unificada (SISU) e o Exame Nacional do Ensino Médio (Enem) (43 - 3%). Há também o ingresso por Mudança de Habilitação (60 - 4%), em que o aluno opta por mudar do curso noturno para o diurno. O ingresso por Transferência Obrigatória (58 – 4%) e as demais formas de ingresso (46 – 3%) também representam uma parcela significativa.

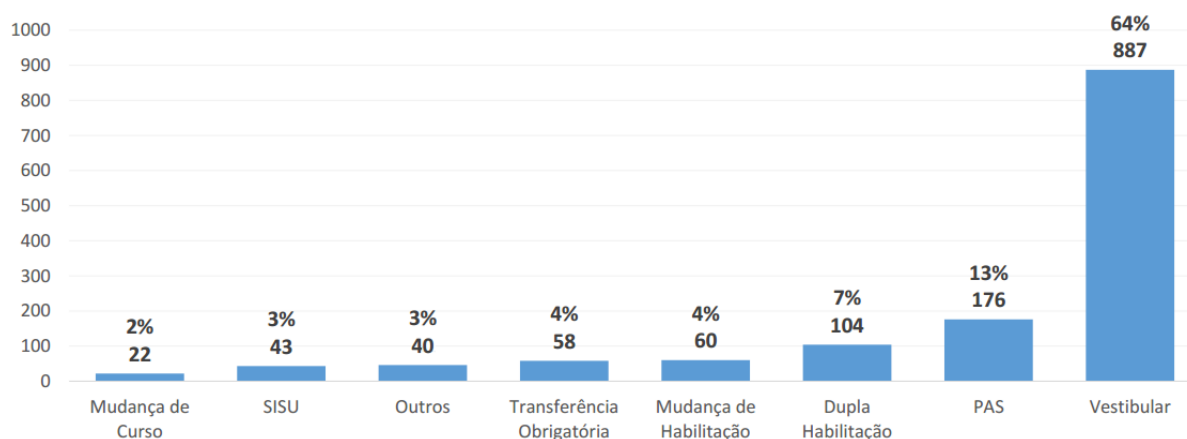


Figura 5.2: Quantidade de alunos ingressados por forma de ingresso (2º/1977 a 1º/2016).

A Figura 5.3 mostra o gráfico do percentual de alunos ingressados pelo sistema universal e por cotas no período de 1º/2005 a 1º/2016. O sistema de cotas foi criado em 2004, e representa 7,9% da forma de ingresso no curso.

De acordo com a Figura 5.4, 940 (68%) alunos que foram ingressados são do gênero masculino, enquanto 450 (32%) são do gênero feminino.

Apesar da grande importância dos temas sobre cotas, gênero e escola de origem, este trabalho teve como foco apenas as análises sobre as altas taxas de reprovações dos alunos nas disciplinas obrigatórias.

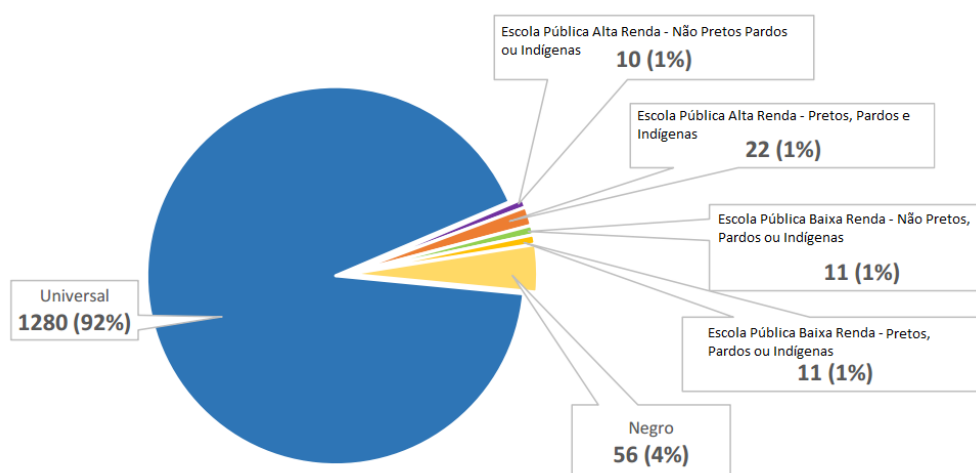


Figura 5.3: Percentual de alunos ingressados pelo sistema universal e por cotas (2º/1977 a 1º/2016).

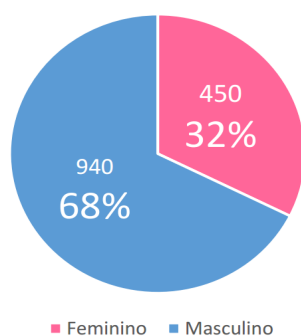


Figura 5.4: Gênero dos alunos ingressados (2º/1977 a 1º/2016).

Na Figura 5.5, 982 (71%) alunos não informaram o tipo de escola de origem, enquanto 220 (16%) vieram de escolas particulares e 188 (14%) vieram de escolas públicas.

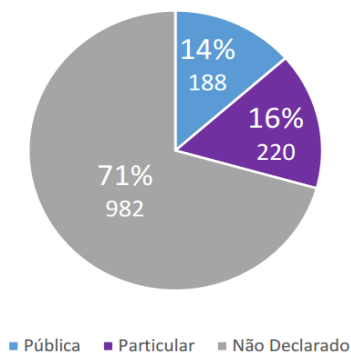


Figura 5.5: Tipo de escola de origem (2º/1977 a 1º/2016).

### 5.3 A Saída do Curso

Ao observar a Figura 5.6, que mostra a quantidade de alunos que saíram do curso por cada forma de saída, podemos ver que a maioria evadiu. A quantidade total de alunos evadidos do curso representa 62% enquanto a quantidade total de formados representa apenas 38%. A saída pela Formatura (464 – 38%) representa uma quantidade inferior à soma das formas de saída que representam a evasão do curso pelo aluno, visto que a quantidade dos alunos que foram desligados por não cumprirem condição é de 291 (24%), dos alunos que abandonaram o curso é de 151 (12%) e dos alunos que desligaram voluntariamente é de 148 (12%). O restante, que representa 15% do total, saiu por um novo vestibular; ou mudança de curso; ou transferência; ou reprovaram 3 vezes a mesma disciplina obrigatória; ou mudaram de habilitação; ou fizeram um novo vestibular para outra habilitação; ou foram desligados por baixo rendimento acadêmico; ou por anulação de registro.

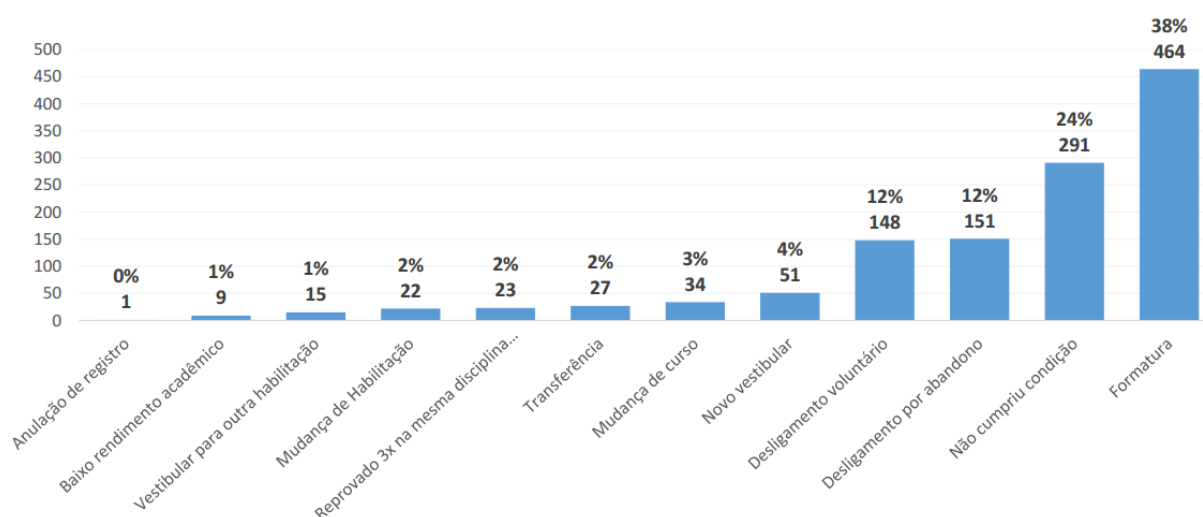


Figura 5.6: Porcentagens de alunos relativas à forma de saída do curso (2º/1977 a 1º/2016).

De acordo com a Análise do Sistema de Cotas para Negros da Universidade de Brasília [14] o sistema de cotas para negros na UnB teve início no vestibular do segundo semestre de 2004. O sistema foi colocado em prática após muito debate e tem como objetivo instaurar no espaço acadêmico um mecanismo reparador das perdas infringidas à população negra brasileira.

A Figura 5.7 mostra a quantidade de alunos formados e evadidos pelo sistema de ingresso. Dos alunos que ingressaram pelo sistema universal, 450 (38%) se formaram, enquanto 723 (62%) evadiram. Os alunos que ingressaram pelo sistema de cotas, 14 (23%)

se formaram, enquanto 47 (77%) evadiram. Pode-se notar que no sistema universal, dos alunos que saíram do curso, mais da metade dos alunos evadiram em relação à quantidade de alunos que se formaram. Enquanto essa diferença no sistema de cotas é de menos de um terço.

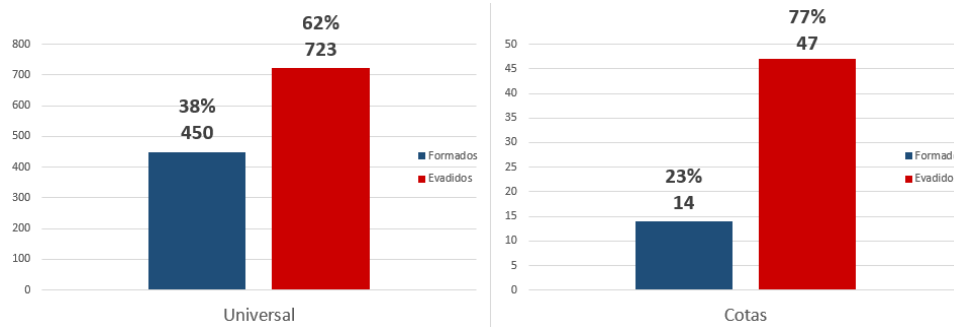


Figura 5.7: A saída do curso por sistema de entrada (2º/1977 a 1º/2016).

## 5.4 Taxas de Reprovação em Disciplinas Obrigatórias

Com o objetivo de verificar quais disciplinas obrigatórias estão mais relacionadas à evasão no curso de Licenciatura em Matemática (Diurno), foram realizadas análises estatísticas sobre a taxa de reprovação em cada disciplina obrigatória de cada semestre do curso, com base no currículo do Projeto Pedagógico do Curso (PPC), vigente em 1º/2014. Para o estudo, foram consideradas como relevantes somente as disciplinas obrigatórias que possuem altas taxas de reprovação (aproximadamente 30% de reprovação).

Apenas as reprovações dos alunos que já saíram do curso, os evadidos e os formados, foram consideradas para a realização das análises estatísticas. As reprovações dos alunos que ainda estão cursando não foram utilizadas, pois não classificam a evasão no curso. Do total de 1.390 alunos ingressados no curso, 159 ainda estão cursando, portanto, iremos considerar apenas as reprovações dos 1.231 alunos que saíram do curso, por formatura ou evasão.

De acordo com o Portal da UnB [12], o sistema de avaliação de desempenho acadêmico considera menções de aprovação e reprovação em cada disciplina concluída pelo aluno.

As menções de reprovação consideradas foram:

- SR: Sem Rendimento, caso o aluno tenha sido reprovado com nota 0 ou abandono da disciplina (acima de 25% de faltas).
- II: Inferior Insuficiente, caso o aluno tenha sido reprovado com nota superior a 0 e inferior a 3.

- MI: Média Inferior, caso o aluno tenha sido reprovado com nota igual ou superior a 3 e inferior a 5.

As menções de aprovação consideradas foram:

- MM: Média Mínima, caso o aluno tenha sido aprovado com nota igual ou superior a 5 e inferior a 7.
- MS: Média Superior, caso o aluno tenha sido aprovado com nota superior ou igual a 7 e inferior a 9.
- SS: Significado Superior, caso o aluno tenha sido aprovado com nota igual ou superior 9.

### 5.4.1 Primeiro Semestre

No primeiro semestre do PPC estão previstas as disciplinas: Cálculo 1 (código: 113034), Geometria Analítica Matemática (código: 105881), Introdução à Ciência da Computação (código: 113913) e Organização da Educação Brasileira (código: 194221). A disciplina Geometria Analítica Matemática foi introduzida no novo PPC a partir de 1º/2014, logo não possui muitos dados para análise estatística. Já a disciplina Organização da Educação Brasileira não possui uma alta taxa de reprovação (12%). Portanto serão consideradas somente as disciplinas Cálculo 1 e Introdução à Ciência da Computação para análise.

#### Cálculo 1

A disciplina Cálculo 1 é ofertada pelo Departamento de Matemática e não possui pré-requisitos. A Figura 5.8 e a Figura 5.9 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

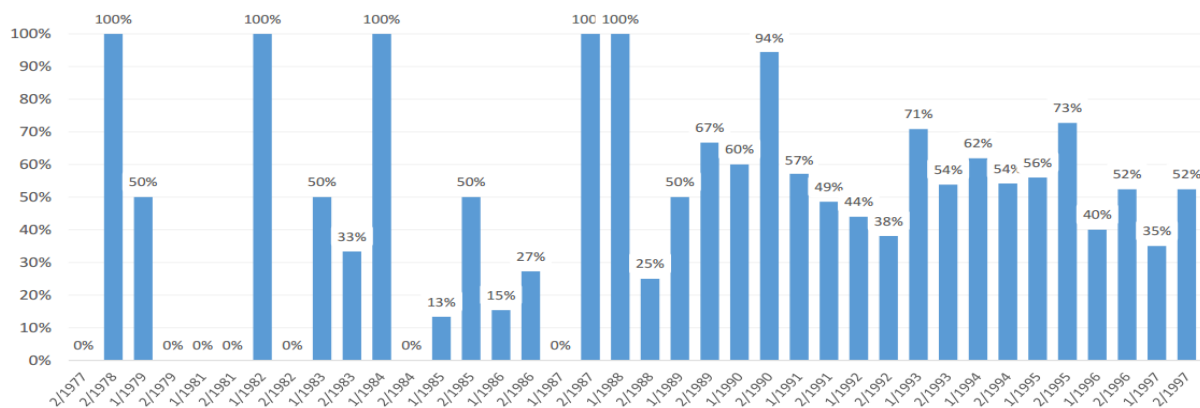


Figura 5.8: Taxa de reprovação em Cálculo 1 no período de 2º/1977 a 2º/1997.

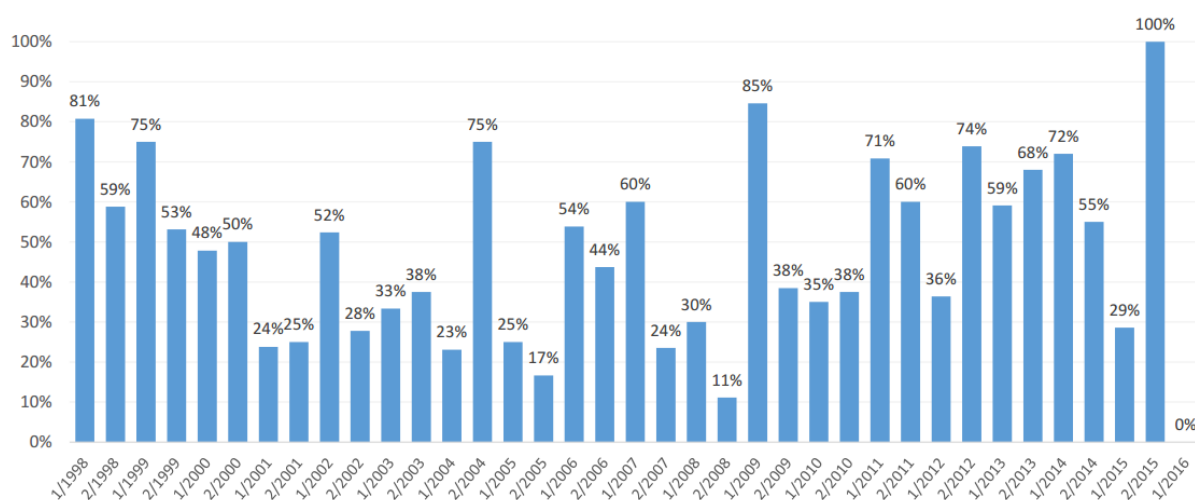


Figura 5.9: Taxa de reprovação em Cálculo 1 no período de 1º/1998 a 1º/2016.

A média de reprovação em Cálculo 1 entre o período de 2º/1977 a 1º/2016 foi de 48%, com um desvio padrão de 7% e variância de 27%. A Figura 5.10 mostra o resultado da análise da disciplina.

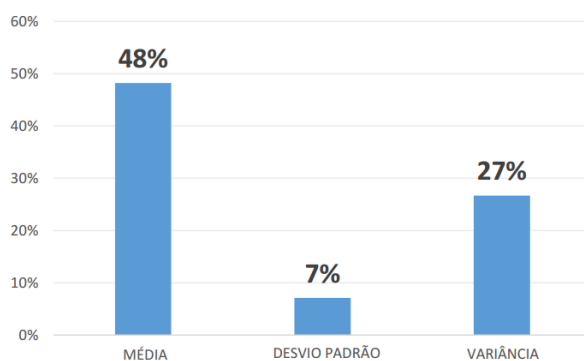


Figura 5.10: Média de reprovação em Cálculo 1 entre o período de 2º/1977 a 1º/2016.

## Introdução à Ciência da Computação

A disciplina Introdução à Ciência da Computação é ofertada pelo Departamento de Ciência da Computação e não possui pré-requisitos. A Figura 5.11 e a Figura 5.12 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

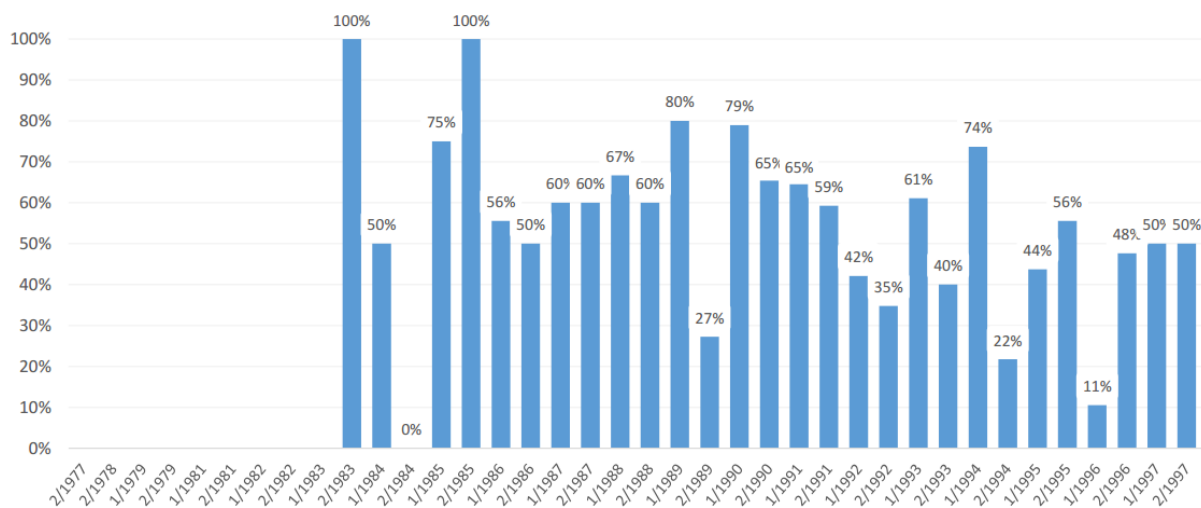


Figura 5.11: Taxa de reprovação em Introdução à Ciência da Computação no período de 2º/1977 a 2º/1997.

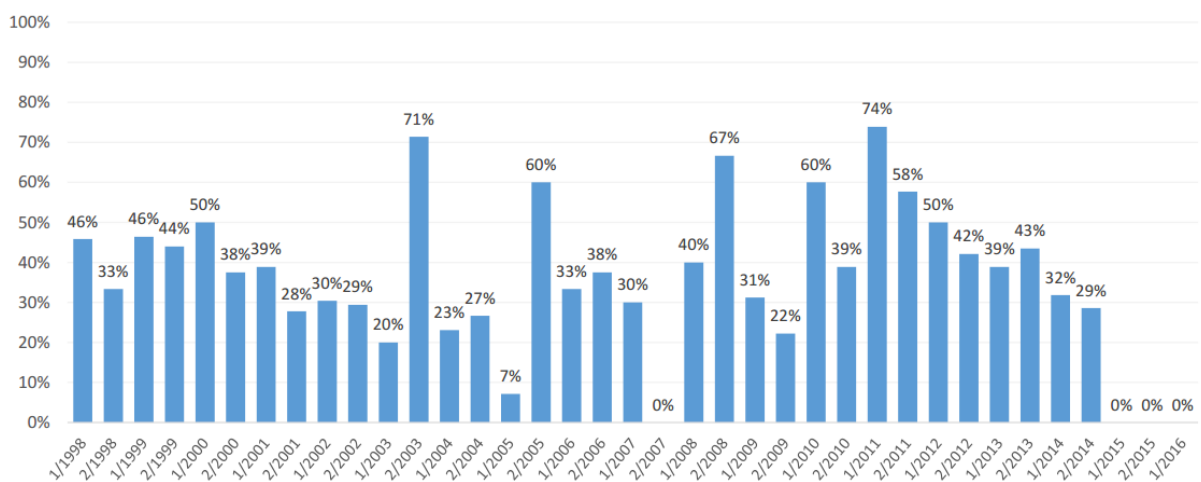


Figura 5.12: Taxa de reprovação em Introdução à Ciência da Computação no período de 1º/1998 a 1º/2016.

A média de reprovação em Introdução à Ciência da Computação entre o período de 2º/1977 a 1º/2016 foi de 45%, com um desvio padrão de 5% e variância de 22%. A Figura 5.13 mostra o resultado da análise da disciplina.



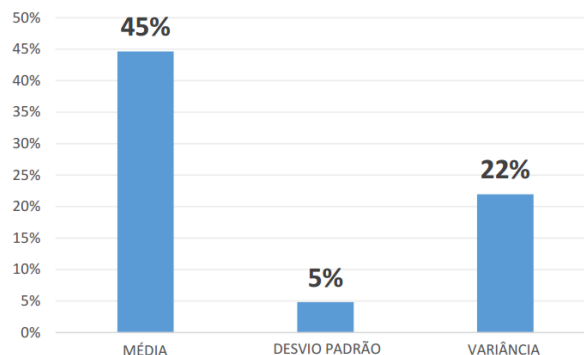


Figura 5.13: Média de reprovação em Introdução à Ciência da Computação entre o período de 2º/1977 a 1º/2016.

## 5.4.2 Segundo Semestre

No segundo semestre estão previstas as disciplinas: Cálculo 2 (código: 113042), Introdução à Álgebra Linear (código: 113093), Desenvolvimento Psicológico e Ensino (código: 125156), Física 1 (código: 118001) e Física 1 Experimental (código: 118010). Não será considerada a disciplina Desenvolvimento Psicológico e Ensino, pois ela não possui uma alta taxa de reprovação (6%).

### Cálculo 2

A disciplina Cálculo 2 é ofertada pelo Departamento de Matemática e possui como pré-requisito a disciplina Cálculo 1. A Figura 5.14 e a Figura 5.15 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

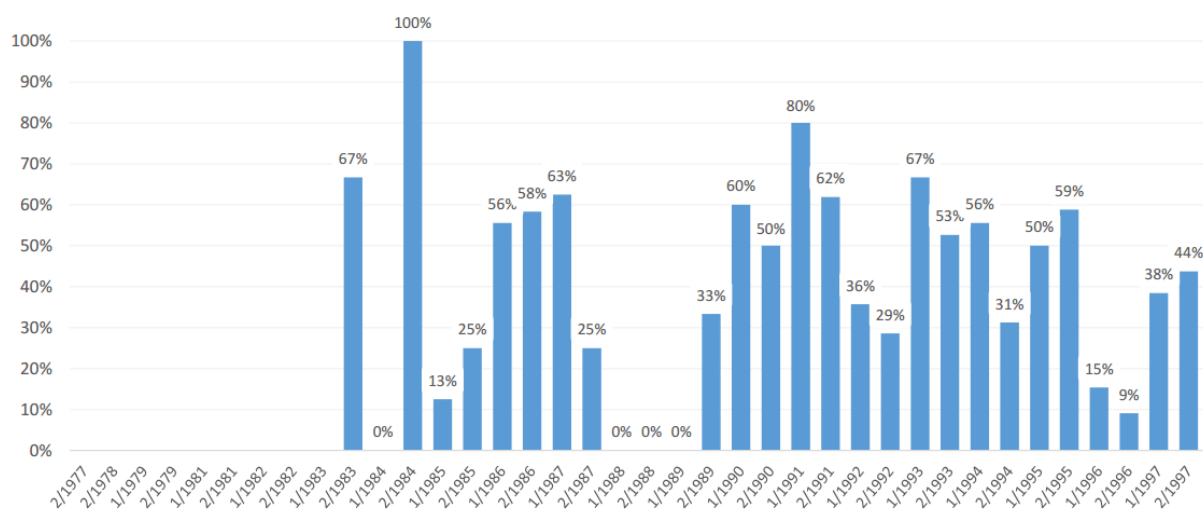


Figura 5.14: Taxa de reprovação em Cálculo 2 no período de 2º/1977 a 2º/1997.

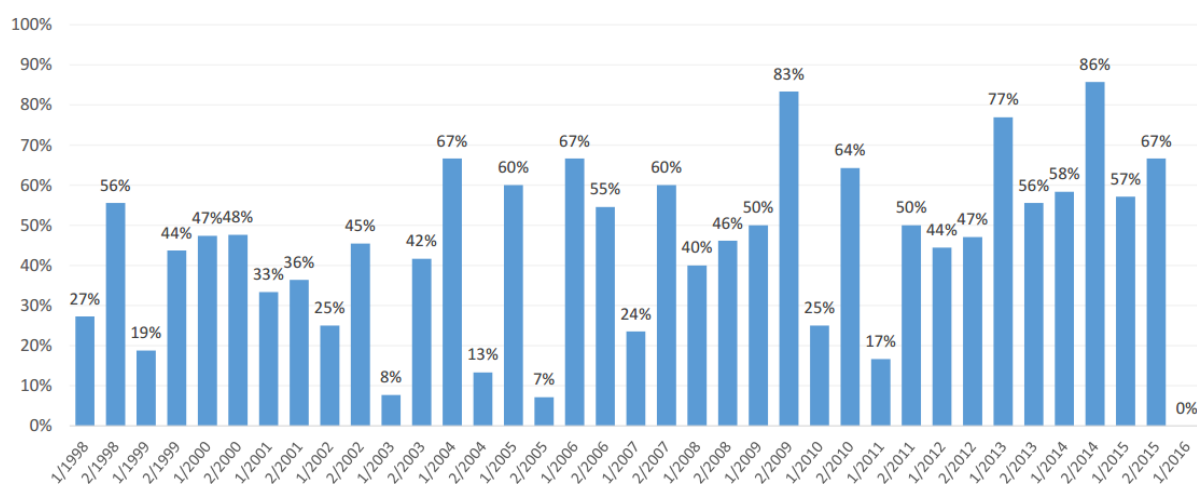


Figura 5.15: Taxa de reprovação em Cálculo 2 no período de 1º/1998 a 1º/2016.

A média de reprovação em Cálculo 2 entre o período de 2º/1977 a 1º/2016 foi de 43%, com um desvio padrão de 5% e variância de 23%. A Figura 5.16 mostra o resultado da análise da disciplina.

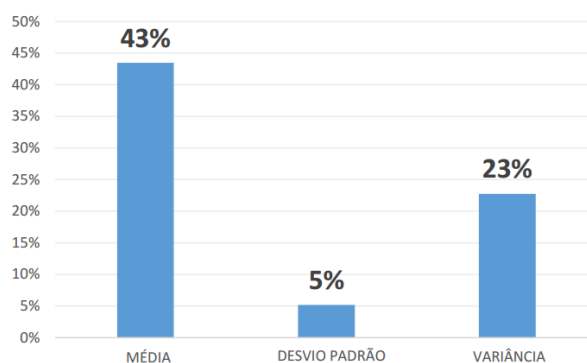


Figura 5.16: Média de reprovação em Cálculo 2 entre o período de 2º/1977 a 1º/2016.

## Introdução à Álgebra Linear

A disciplina Introdução à Álgebra Linear é ofertada pelo Departamento de Matemática e não possui pré-requisitos. A Figura 5.17 e a Figura 5.18 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

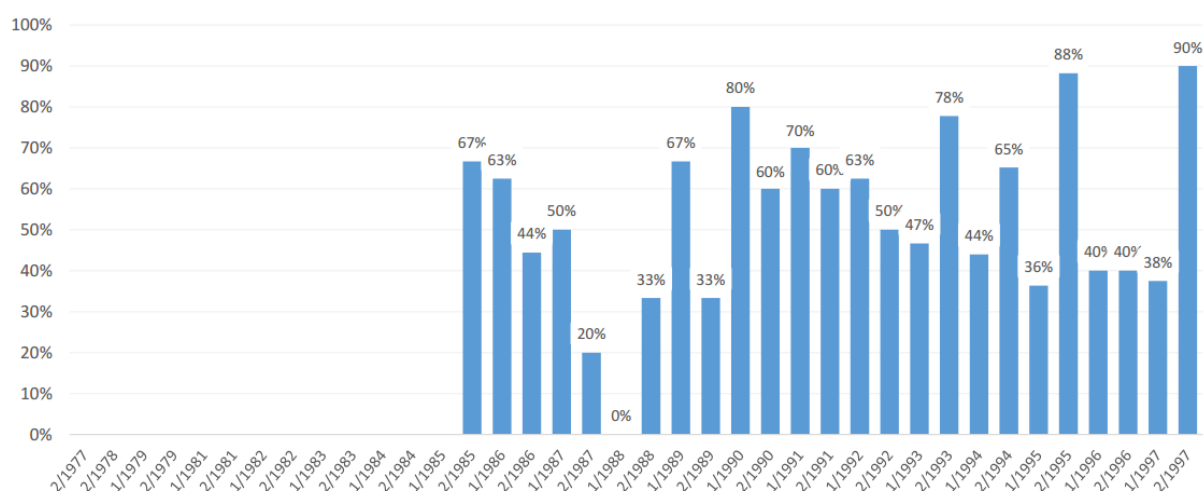


Figura 5.17: Taxa de reprovação em Introdução à Álgebra Linear no período de 2º/1977 a 2º/1997.

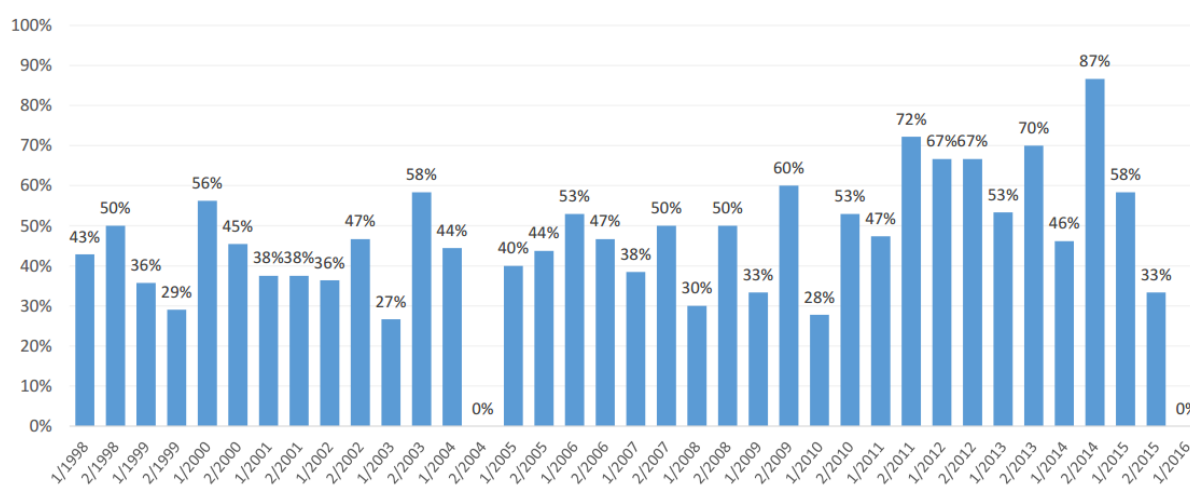


Figura 5.18: Taxa de reprovação em Introdução à Álgebra Linear no período de 1º/1998 a 1º/2016.

A média de reprovação em Introdução à Álgebra Linear entre o período de 2º/1977 a 1º/2016 foi de 46%, com um desvio padrão de 4% e variância de 21%. A Figura 5.19 mostra o resultado da análise da disciplina.

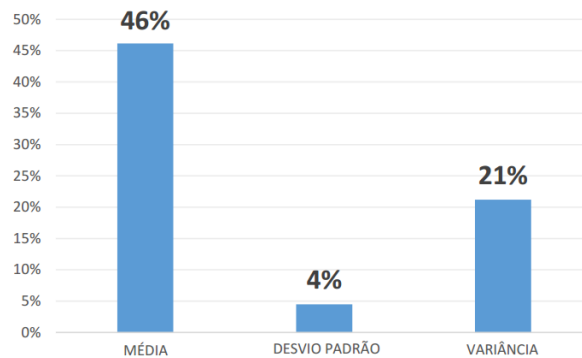


Figura 5.19: Média de reprovação em Introdução à Álgebra Linear entre o período de 2º/1977 a 1º/2016.

## Física 1

A disciplina Física 1 é ofertada pelo Instituto de Física e não possui pré-requisitos. A Figura 5.20 e a Figura 5.21 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

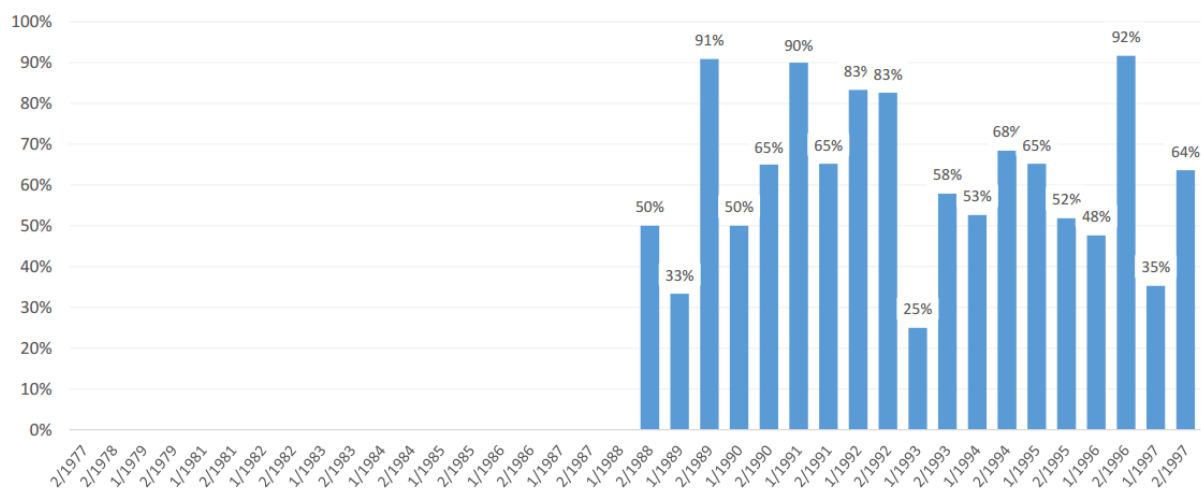


Figura 5.20: Taxa de reprovação em Física 1 no período de 2º/1977 a 2º/1997.

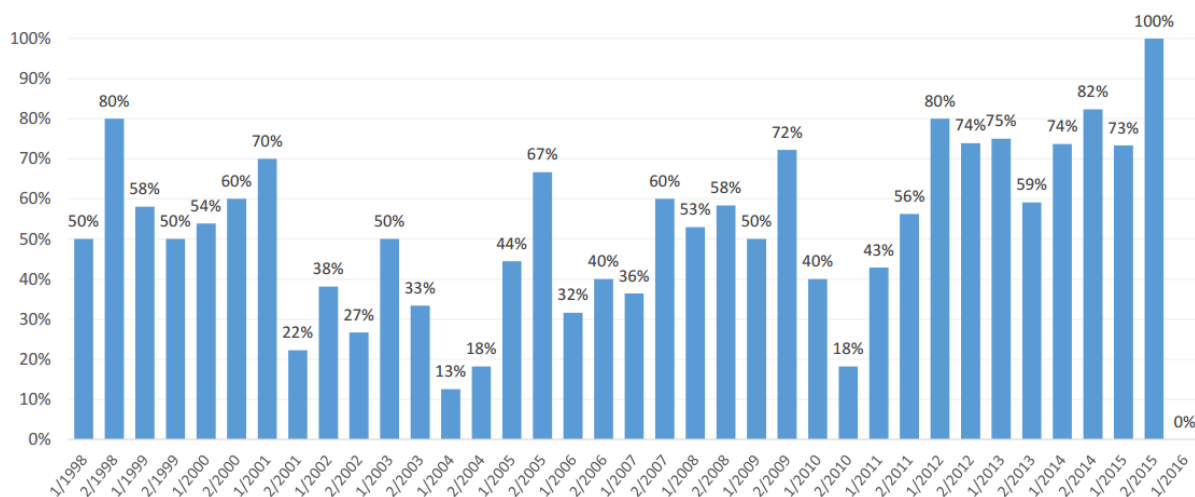


Figura 5.21: Taxa de reprovação em Física 1 no período de 1º/1998 a 1º/2016.

A média de reprovação em Física 1 entre o período de 2º/1977 a 1º/2016 foi de 55%, com um desvio padrão de 5% e variância de 22%. A Figura 5.22 mostra o resultado da análise da disciplina.

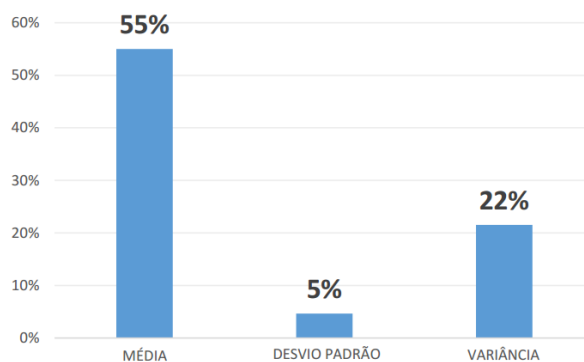


Figura 5.22: Média de reprovação em Física 1 entre o período de 2º/1977 a 1º/2016.

## Física 1 Experimental

A disciplina Física 1 Experimental é ofertada pelo Instituto de Física e não possui pré-requisitos. A Figura 5.23 e a Figura 5.24 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

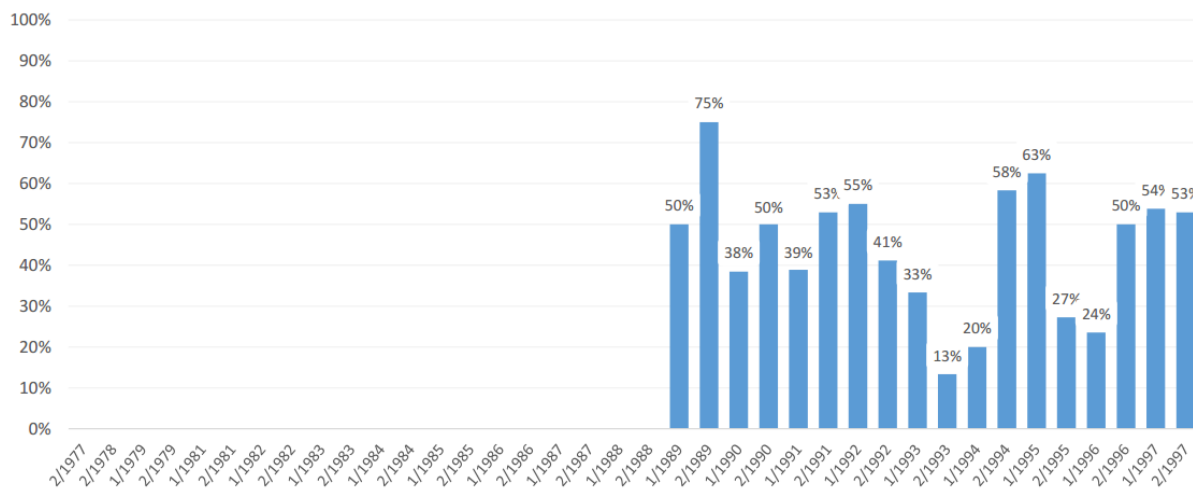


Figura 5.23: Taxa de reprovação em Física 1 Experimental no período de 2º/1977 a 2º/1997.

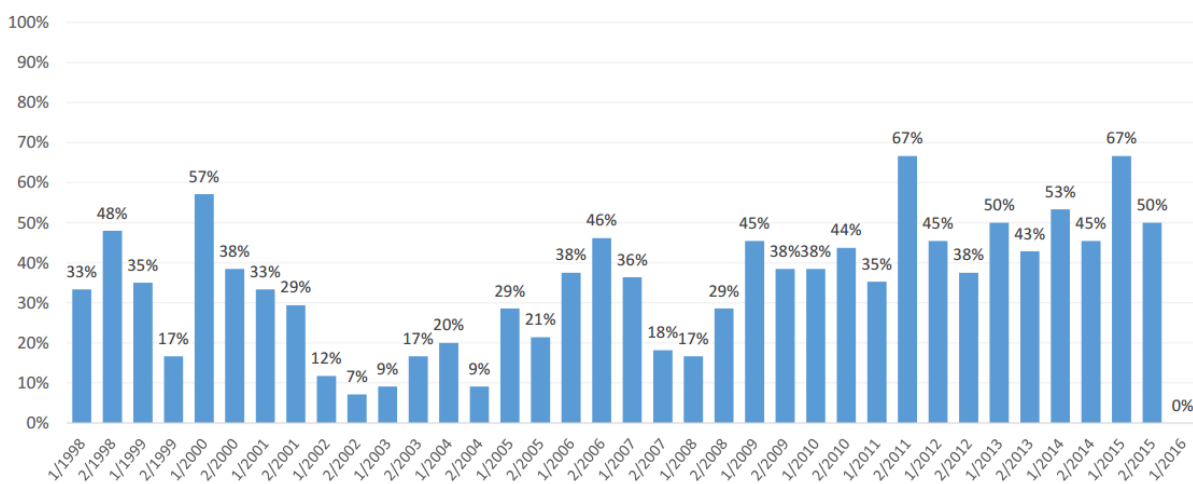


Figura 5.24: Taxa de reprovação em Física 1 Experimental no período de 1º/1998 a 1º/2016.

A média de reprovação em Física 1 Experimental entre o período de 2º/1977 a 1º/2016 foi de 55%, com um desvio padrão de 5% e variância de 22%. A Figura 5.25 mostra o resultado da análise da disciplina.

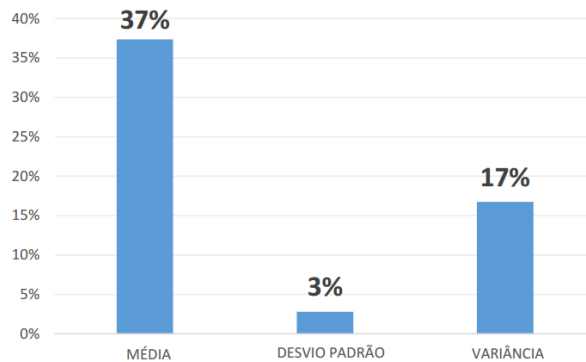


Figura 5.25: Média de reprovação em Física 1 Experimental entre o período de 2º/1977 a 1º/2016.

### 5.4.3 Terceiro Semestre

No terceiro semestre estão previstas as disciplinas: Cálculo 3 (código: 113051), Teoria dos Números (código: 113115) e Física 2 (código: 118028). Todas disciplinas possuem altos índices de reprovação.

#### Cálculo 3

A disciplina Cálculo 3 é ofertada pelo Departamento de Matemática e possui como pré-requisito a disciplina Cálculo 2. A Figura 5.26 e a Figura 5.27 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

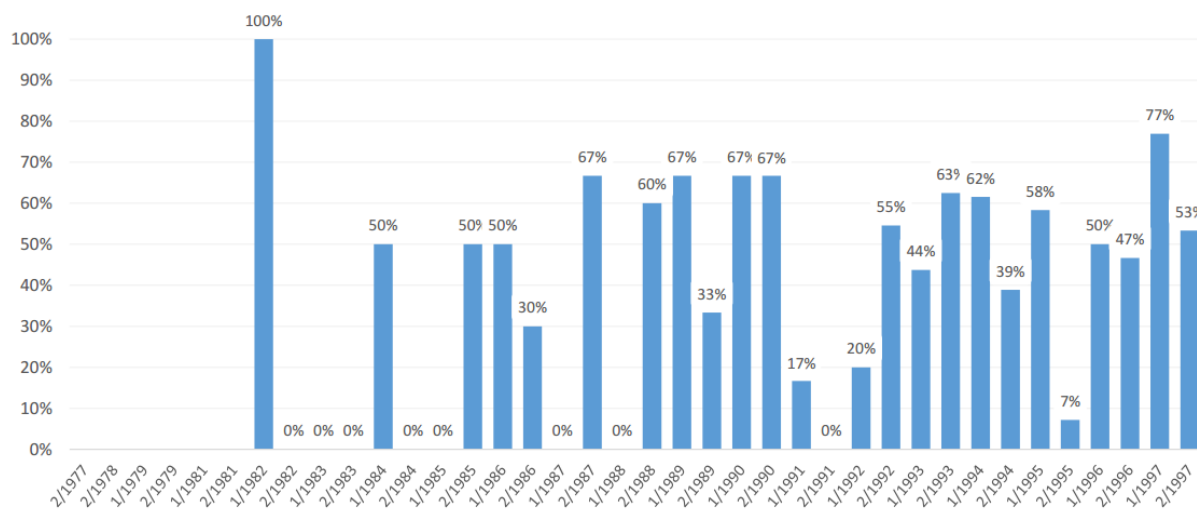


Figura 5.26: Taxa de reprovação em Cálculo 3 no período de 2º/1977 a 2º/1997.

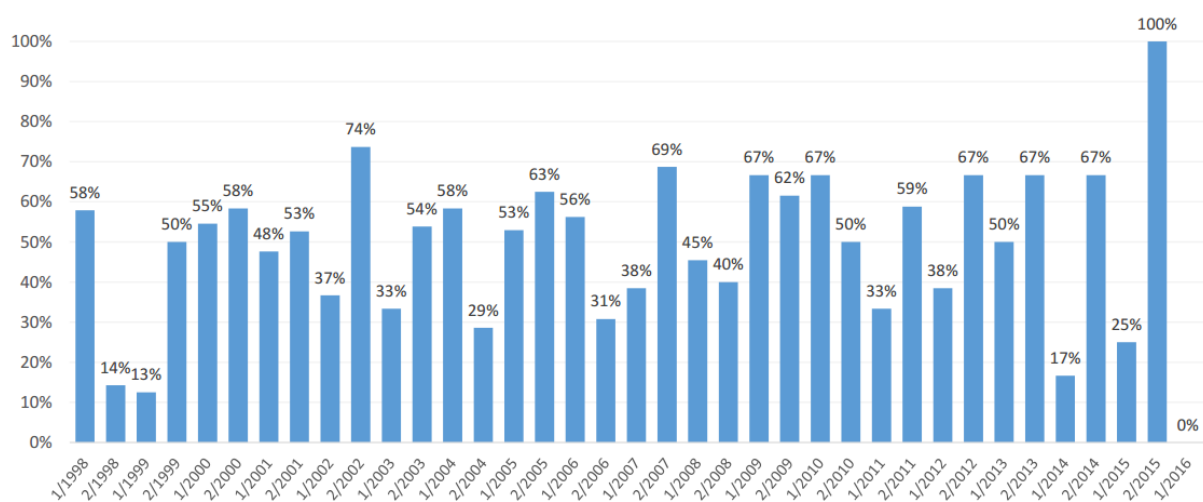


Figura 5.27: Taxa de reprovação em Cálculo 3 no período de 1º/1998 a 1º/2016.

A média de reprovação em Cálculo 3 entre o período de 2º/1977 a 1º/2016 foi de 46%, com um desvio padrão de 5% e variância de 23%. A Figura 5.28 mostra o resultado da análise da disciplina.

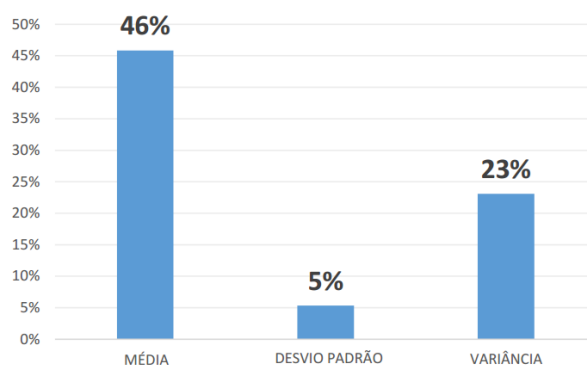


Figura 5.28: Média de reprovação em Cálculo 3 entre o período de 2º/1977 a 1º/2016.



## Teoria dos Números

A disciplina Teoria dos Números é ofertada pelo Departamento de Matemática e não possui pré-requisitos. A Figura 5.29 e a Figura 5.30 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

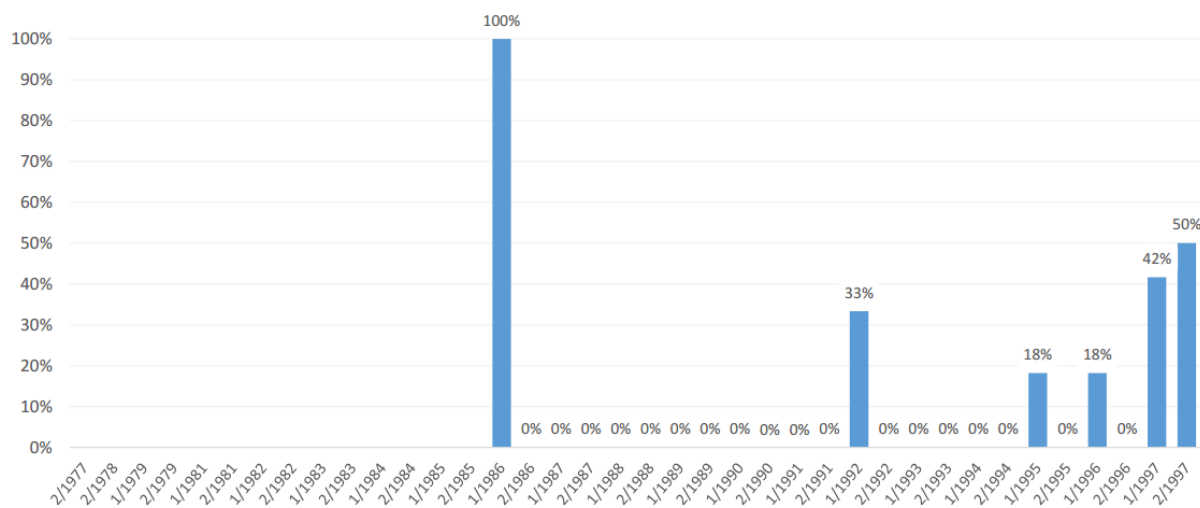


Figura 5.29: Taxa de reprovação em Teoria dos Números no período de 2º/1977 a 2º/1997.

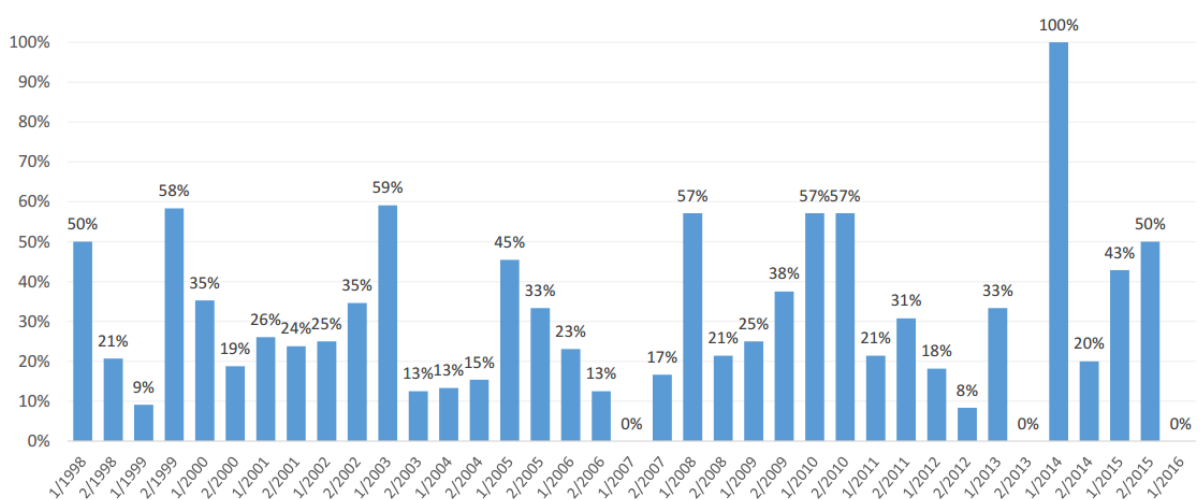


Figura 5.30: Taxa de reprovação em Teoria dos Números no período de 1º/1998 a 1º/2016.

A média de reprovação em Teoria dos Números entre o período de 2º/1977 a 1º/2016 foi de 28%, com um desvio padrão de 5% e variância de 23%. A Figura 5.31 mostra o resultado da análise da disciplina.

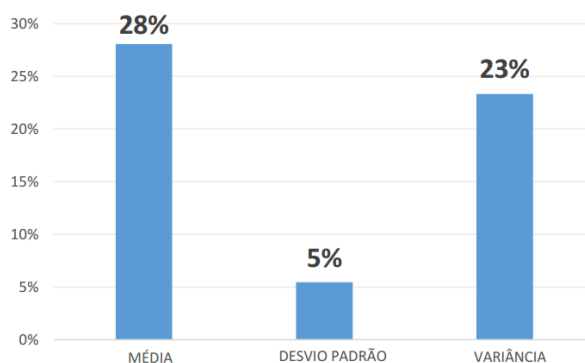


Figura 5.31: Média de reprovação em Teoria dos Números entre o período de 2º/1977 a 1º/2016.

## Física 2

A disciplina Física 2 é ofertada pelo Instituto de Física e possui como pré-requisitos Cálculo 1 e Física 1. A Figura 5.32 e a Figura 5.33 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016. Entre o período de 2º/1977 a 1º/1988 não houve matrículas nessa disciplina.

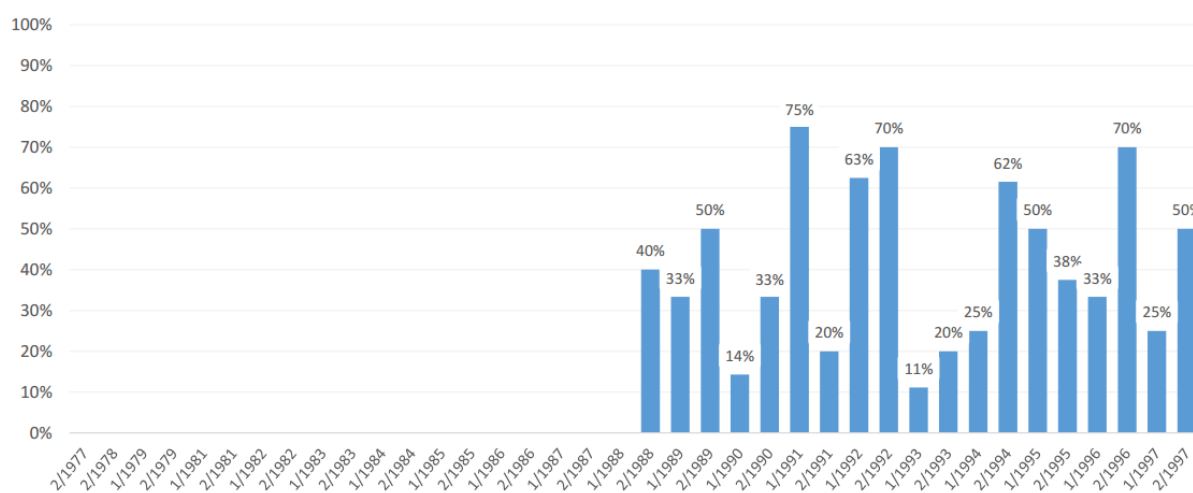


Figura 5.32: Taxa de reprovação em Física 2 no período de 2º/1977 a 2º/1997.

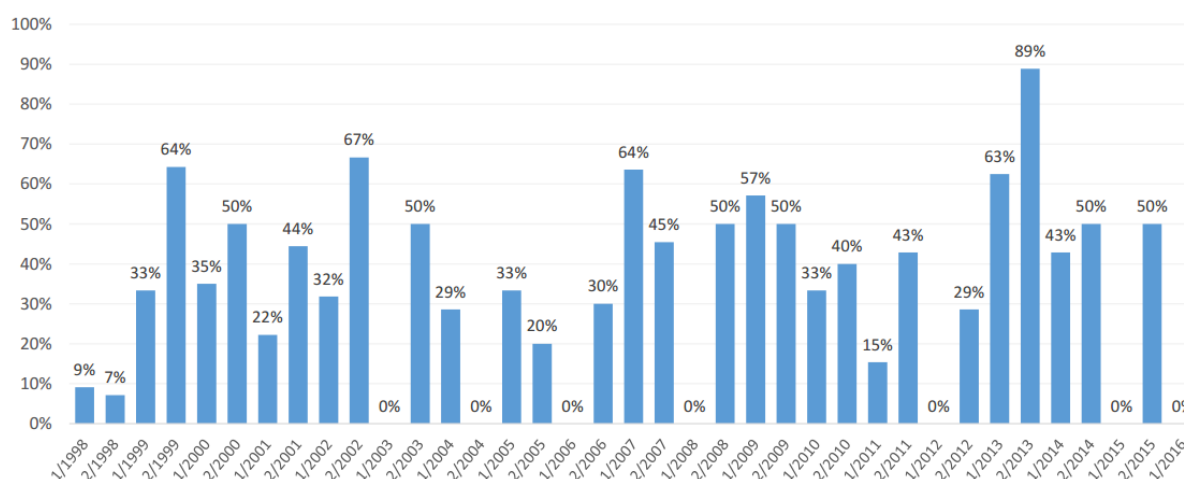


Figura 5.33: Taxa de reprovação em Física 2 no período de 1º/1998 a 1º/2016.

A média de reprovação em Física 2 entre o período de 2º/1977 a 1º/2016 foi de 38%, com um desvio padrão de 5% e variância de 21%. A Figura 5.34 mostra o resultado da análise da disciplina.

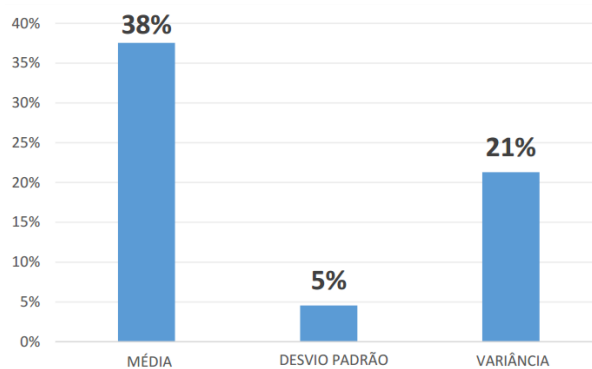


Figura 5.34: Média de reprovação em Física 2 entre o período de 2º/1977 a 1º/2016.

#### 5.4.4 Quarto Semestre

No quarto semestre estão previstas as disciplinas: Álgebra 1 (código: 113107), Geometria 1 (código: 117161), Aprendizagem no Ensino (código: 125172), Didática Fundamental (código: 192015) e Direitos Humanos e Cidadania (código: 199851). A única disciplina que será relevante para o estudo é Álgebra 1. As demais disciplinas serão desconsideradas, por não ter um alto índice de reprovação, como Geometria 1 (22%) e Didática Fundamental (11%); Por não ser obrigatória, como Direitos Humanos e Cidadania; E por ter sido introduzida no novo PPC a partir de 1º/2014 e não possuir muitos dados, como Aprendizagem no Ensino.

#### Álgebra 1

A disciplina Álgebra 1 é ofertada pelo Departamento de Matemática e não possui pré-requisitos. A Figura 5.35 e a Figura 5.27 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

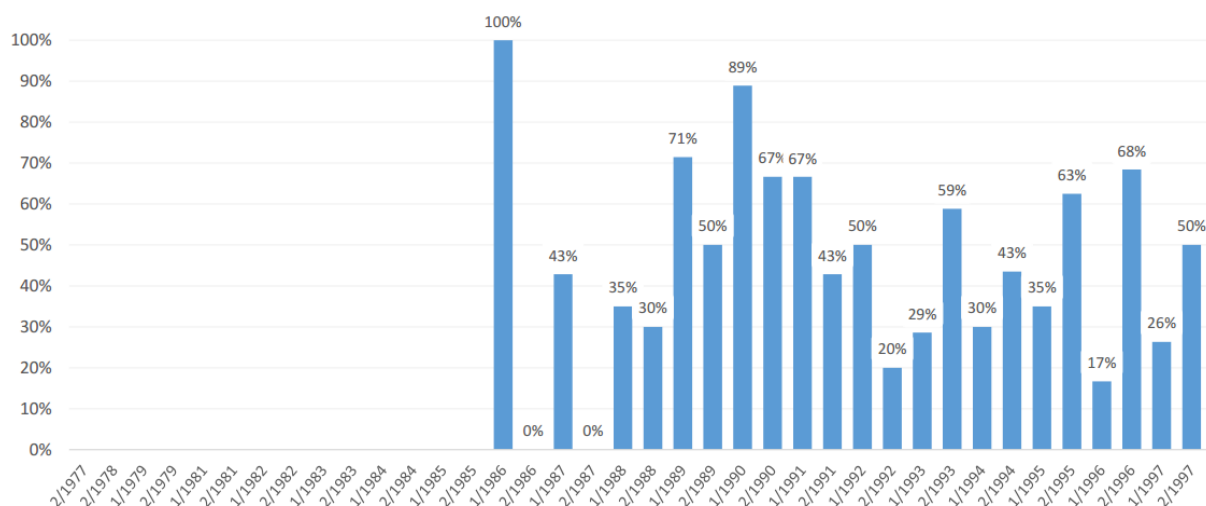


Figura 5.35: Taxa de reprovação em Álgebra 1 no período de 2º/1977 a 2º/1997.

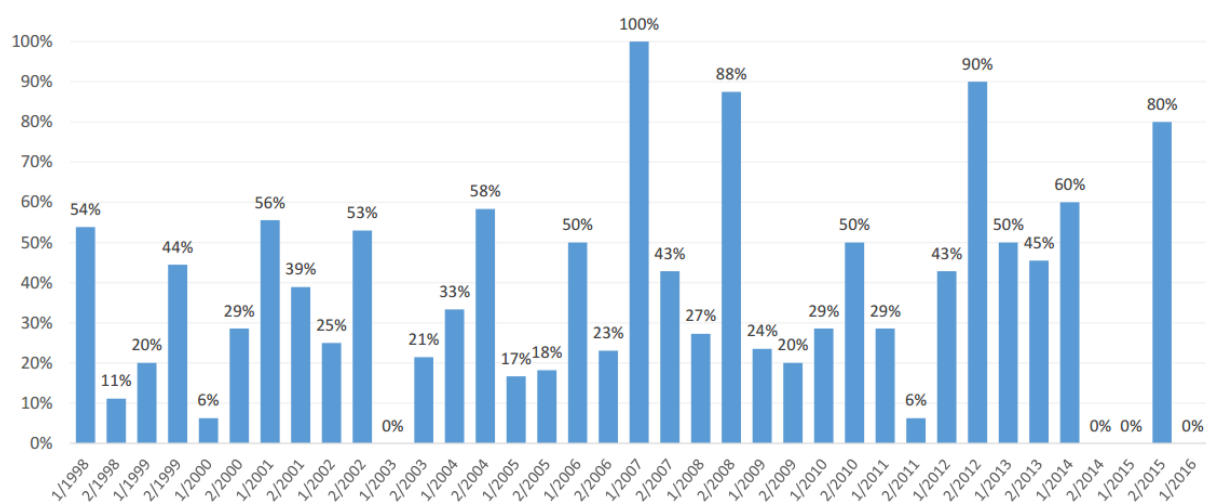


Figura 5.36: Taxa de reprovação em Álgebra 1 no período de 1º/1998 a 1º/2016.

A média de reprovação em Álgebra 1 entre o período de 2º/1977 a 1º/2016 foi de 41%, com um desvio padrão de 6% e variância de 25%. A Figura 5.37 mostra o resultado da análise da disciplina.

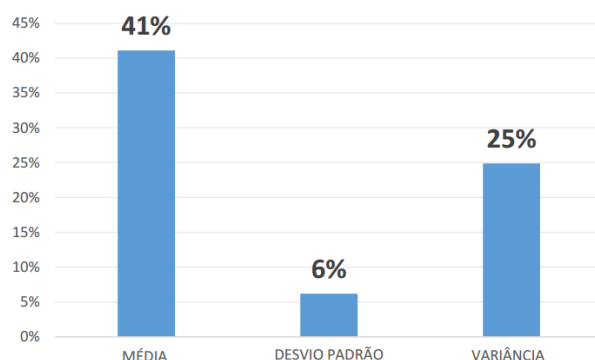


Figura 5.37: Média de reprovação em Álgebra 1 entre o período de 2º/1977 a 1º/2016.

#### 5.4.5 Quinto Semestre

No quinto semestre estão previstas as disciplinas: Cálculo de Probabilidade 1 (código: 113824) e Geometria 2 (código: 117170). Nenhuma disciplina será relevante para o estudo, pois Cálculo de Probabilidade 1 e Geometria 2 foram introduzidas apenas no novo PPC a partir de 1º/2014.

### 5.4.6 Sexto Semestre

No sexto semestre estão previstas as disciplinas: Álgebra para o Ensino 1 (código: 117421), Análise 1 (código: 113204) e Geometria para o Ensino 1 (código: 117471). Apenas a disciplina Análise 1 será relevante para o estudo visto que as disciplinas Álgebra para o Ensino 1 e Geometria para o Ensino 1 foram introduzidas no novo PPC a partir de 1º/2014.

#### Análise 1

A disciplina Análise 1 é ofertada pelo Departamento de Matemática e possui como pré-requisito a disciplina Cálculo 2. A Figura 5.38 e a Figura 5.39 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

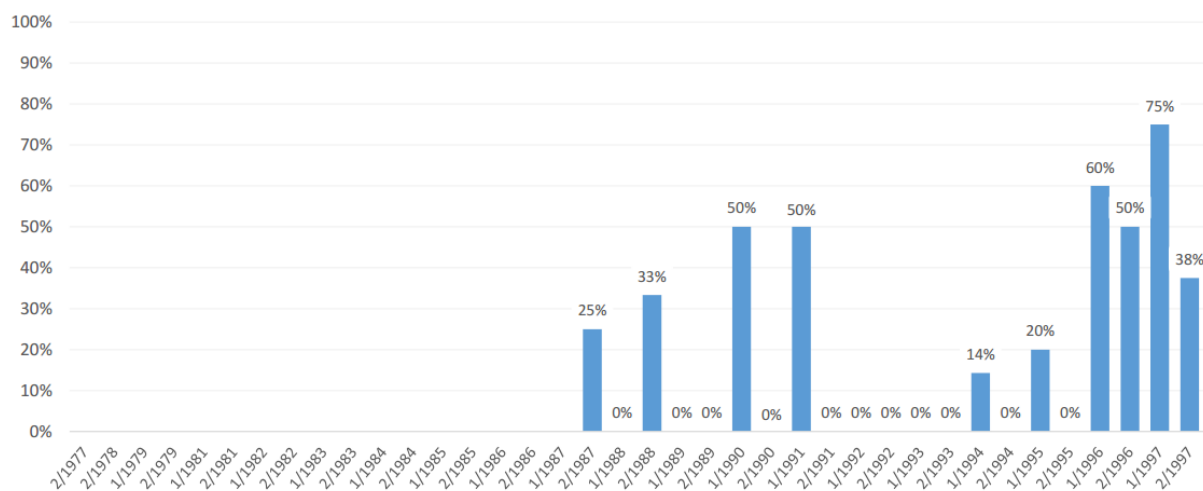


Figura 5.38: Taxa de reprovação em Análise 1 no período de 2º/1977 a 2º/1997.

A média de reprovação em Análise 1 entre o período de 2º/1977 a 1º/2016 foi de 36%, com um desvio padrão de 7% e variância de 27%. A Figura 5.40 mostra o resultado da análise da disciplina.

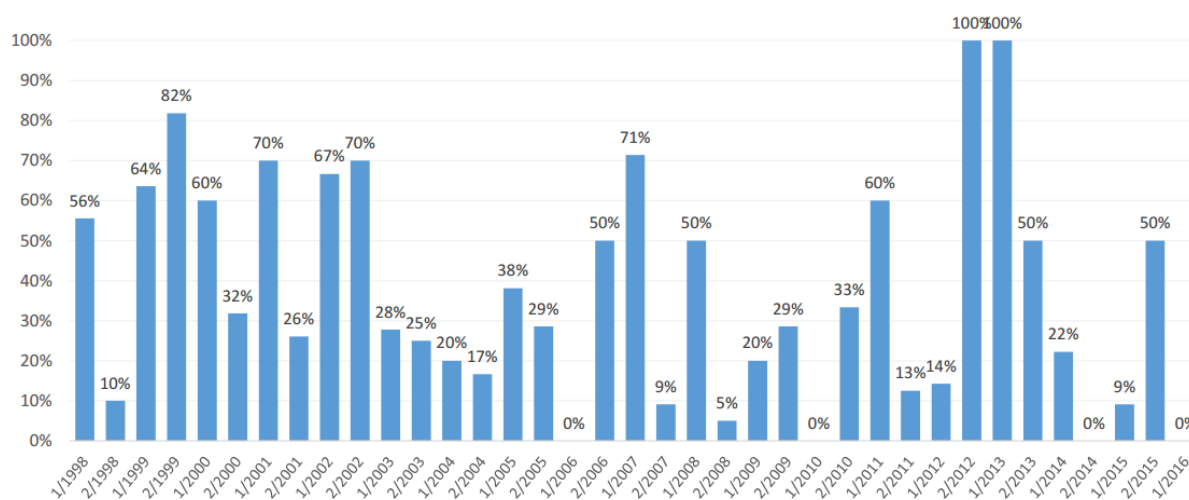


Figura 5.39: Taxa de reprovação em Análise 1 no período de 1º/1998 a 1º/2016.

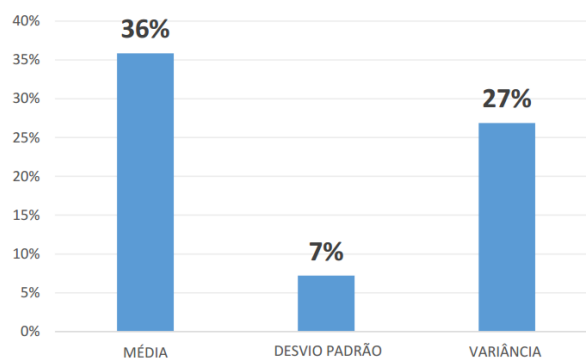


Figura 5.40: Média de reprovação em Análise 1 entre o período de 2º/1977 a 1º/2016.

### 5.4.7 Sétimo Semestre

No sétimo semestre estão previstas as disciplinas: Regência 2 (código: 117439), Geometria para o Ensino 2 (código: 117480) e Álgebra para o Ensino 2 (código: 117501). Nenhuma disciplina será relevante para o estudo, pois todas disciplinas foram introduzidas no novo PPC a partir de 1º/2014.

### 5.4.8 Oitavo Semestre

No oitavo semestre estão previstas as disciplinas: Variável Complexa 1 (código: 113069) e Regência 1 (código: 117510). Apenas a disciplina Variável Complexa 1 será considerada para o estudo, pois Regência 1 foi introduzida no novo PPC a partir de 1º/2014.

## Variável Complexa 1

A disciplina Variável Complexa 1 é ofertada pelo Departamento de Matemática e possui como pré-requisito a disciplina Cálculo 3. A Figura 5.41 e a Figura 5.42 mostram as taxas de reprovação na disciplina entre o período de 2º/1977 a 1º/2016.

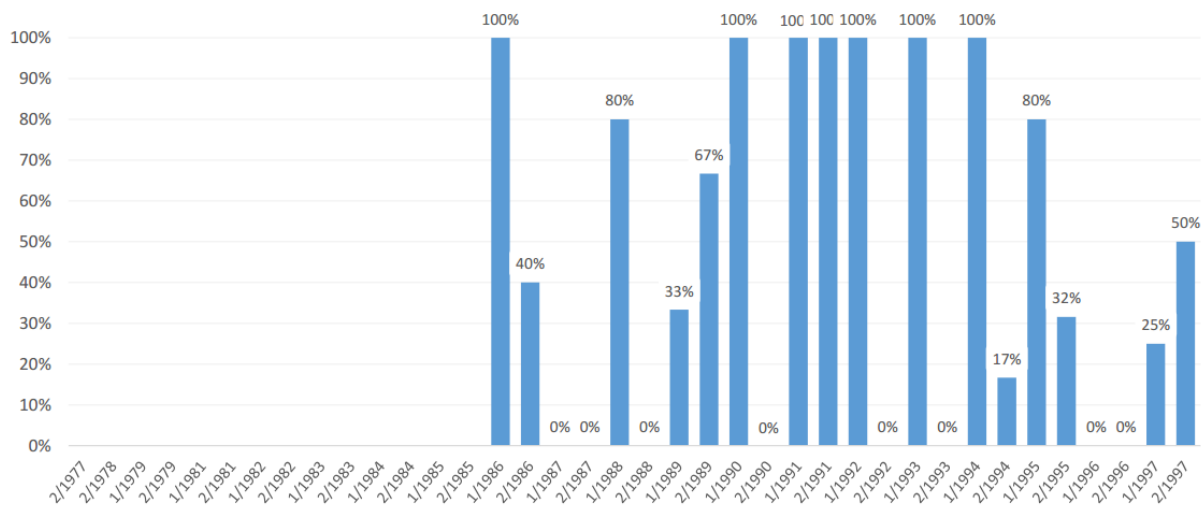


Figura 5.41: Taxa de reprovação em Variável Complexa 1 no período de 2º/1977 a 2º/1997.

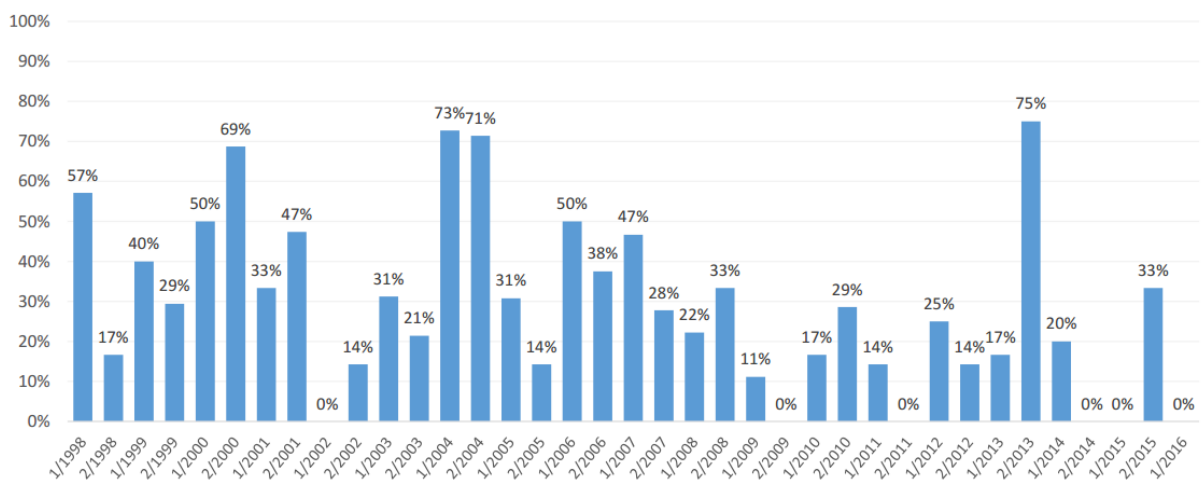


Figura 5.42: Taxa de reprovação em Variável Complexa 1 no período de 1º/1998 a 1º/2016.



A média de reprovação em Variável Complexa 1 entre o período de 2º/1977 a 1º/2016 foi de 35%, com um desvio padrão de 10% e variância de 32%. A Figura 5.43 mostra o resultado da análise da disciplina.

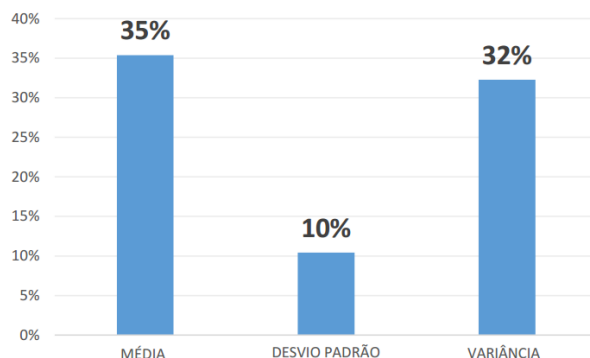


Figura 5.43: Média de reprovação em Variável Complexa 1 entre o período de 2º/1977 a 1º/2016.

## 5.5 Classificação das Disciplinas por Reprovação

Após a realização das análises estatísticas de cada disciplina obrigatória, foi possível obter a classificação de cada disciplina obrigatória por taxa média de reprovação. A Tabela 5.1 mostra a classificação das disciplinas que os alunos mais reprovaram.

Tabela 5.1: Classificação das disciplinas que os alunos mais reprovaram.

Semestre	Código	Nome	Média	Desvio Padrão	Variância
2	118001	Física 1	55%	5%	22%
1	113034	Cálculo 1	48%	7%	27%
3	113051	Cálculo 3	46%	5%	23%
2	113093	Intr. à Álgebra Linear	46%	4%	21%
1	113913	Intr. à Ciência da Computação	45%	5%	22%
2	113042	Cálculo 2	43%	5%	23%
5	113107	Álgebra 1	41%	6%	25%
3	118028	Física 2	38%	5%	21%
2	118010	Física 1 Experimental	37%	3%	17%
6	113204	Análise 1	36%	7%	27%
7	113069	Variável Complexa 1	35%	10%	32%
4	113115	Teoria dos Números	28%	5%	23%

## Capítulo 6

# Disciplinas Obrigatórias Relacionadas à Evasão

Este capítulo apresenta os resultados obtidos pelo processo de mineração de dados no WEKA, que foi realizado para identificar as disciplinas obrigatórias relacionadas à evasão no curso de Licenciatura em Matemática (Diurno) da UnB. Inicialmente, a Seção 6.1 apresenta o processo de mineração de dados, que foi realizado após a identificação das disciplinas obrigatórias que mais reprovam. Em seguida a Seção 6.2 apresenta as conclusões sobre os resultados obtidos.

### 6.1 Mineração de Dados

O objetivo da mineração de dados sobre as disciplinas obrigatórias foi verificar quais conjuntos de disciplinas estão mais relacionados à evasão no curso de Licenciatura em Matemática (Diurno). Para isso, foram selecionadas para a mineração de dados somente as disciplinas obrigatórias que possuíam altas taxas de reprovação, com uma taxa mínima aproximada de 30% de reprovação, pois as disciplinas com baixas taxas de reprovação não fazem boas predições.

Para a realização da mineração de dados sobre as reprovações em disciplinas obrigatórias, foi necessário popular o banco de dados com as menções dos 1.231 alunos em cada disciplina obrigatória, desde o período da criação do curso em 2º/1977 até o semestre 1º/2016. Foram consideradas até 3 tentativas de aprovação nas disciplinas, portanto, para cada disciplina foi criado 3 atributos de tentativas.

Adicionalmente, com o intuito de verificar se existem outras variáveis que estão relacionadas à evasão, além da variável reprovação em disciplinas obrigatórias, os seguintes atributos foram incluídos no banco de dados: SEXO, NASCIMENTO, UF, FORMA INGRESSO OPCA, RACA, ANO INGRESSO OPCA, SEMESTRE INGRESSO OP-

CAO, TIPO ESCOLA, FOR SAIDA OPCA0, ANO SAIDA OPCA0, SEMESTRE SAIDA OPCA0 e IRA.

Como apresentado na Seção 3.4, um método para descoberta de padrões e predição é a classificação. Para realizar o método de classificação, o algoritmo utilizado foi o C4.5. Na ferramenta WEKA, uma implementação na linguagem Java do algoritmo C4.5 é chamada de J48. O algoritmo J48 está disponível para uso no WEKA Explorer, na seção *trees/J48*.

Primeiramente, um experimento de mineração de dados foi realizado ao retirar o atributo MATRICALUNO, pois o algoritmo J48 do WEKA não aceita atributos do tipo cadeia de caracteres (*string*). O resultado do algoritmo J48 é mostrado na Figura 6.1.

```
Test mode:      10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

FOR_SAIDA_OPCA0 = Mudanca_de_Curso: Evadiu (35.0)
FOR_SAIDA_OPCA0 = Vestibular_p/outra_Habilitacao: Evadiu (15.0)
FOR_SAIDA_OPCA0 = Desligamento_Rendimento_Academico: Evadiu (9.0)
FOR_SAIDA_OPCA0 = Formatura: Formou (460.0)
FOR_SAIDA_OPCA0 = Transferencia: Evadiu (26.0)
FOR_SAIDA_OPCA0 = Desligamento_Abandono: Evadiu (150.0)
FOR_SAIDA_OPCA0 = Desligamento_Voluntario: Evadiu (148.0)
FOR_SAIDA_OPCA0 = Deslig_Nao_Cumpriu_condicao: Evadiu (291.0)
FOR_SAIDA_OPCA0 = Mudanca_de_Habilitacao: Evadiu (22.0)
FOR_SAIDA_OPCA0 = Falecimento: Evadiu (0.0)
FOR_SAIDA_OPCA0 = Rep_3_vezes_na_mesma_Disc_Obrig: Evadiu (23.0)
FOR_SAIDA_OPCA0 = Novo_Vestibular: Evadiu (51.0)
FOR_SAIDA_OPCA0 = Anulacao_de_Registro: Evadiu (1.0)
FOR_SAIDA_OPCA0 = Esta_Cursando: Evadiu (0.0)

Number of Leaves   :    14

Size of the tree   :    15

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1231           100      %
Incorrectly Classified Instances      0             0      %
Kappa statistic                      1
Mean absolute error                   0.0003
Root mean squared error               0.0106
Relative absolute error               0.0648 %
Root relative squared error          2.2013 %
Total Number of Instances           1231
```

Figura 6.1: Resultado obtido na mineração de dados com o algoritmo J48, sem o atributo MATRICALUNO.

Analisando o resultado obtido na Figura 6.1, é possível observar que há um viés sobre o atributo FOR SAIDA OPCA0, pois o algoritmo indica que se o aluno saiu do curso por alguma forma de saída relacionada à evasão, então o aluno evadiu. Logo foi preciso retirar o atributo do processo de mineração de dados.

```

Test mode:      10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

IRA <= 2.5257: Evadiu (541.11/39.18)
IRA > 2.5257
|  ANO_SAIDA_OPCA0 <= 1995: Evadiu (163.84/49.34)
|  ANO_SAIDA_OPCA0 > 1995
|  |  NASCIMENTO <= 1993: Formou (493.04/126.56)
|  |  NASCIMENTO > 1993
|  |  |  ANO_INGRESSO_OPCA0 <= 2012
|  |  |  |  ANO_SAIDA_OPCA0 <= 2014: Evadiu (5.0)
|  |  |  |  ANO_SAIDA_OPCA0 > 2014: Formou (4.0)
|  |  |  ANO_INGRESSO_OPCA0 > 2012: Evadiu (24.0/1.0)

Number of Leaves   :      6

Size of the tree   :     11

Time taken to build model: 0.14 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      998           81.0723 %
Incorrectly Classified Instances    233           18.9277 %
Kappa statistic                    0.6024
Mean absolute error                 0.2764
Root mean squared error             0.3737
Relative absolute error             59.0503 %
Root relative squared error         77.2387 %
Total Number of Instances          1231

```

Figura 6.2: Resultado obtido na mineração de dados com o algoritmo J48, sem os atributos MATRICALUNO e FOR SAIDA OPCA0.

Após a retirada do atributo FOR SAIDA OPCA0, o algoritmo classificador J48 mostrou o resultado apresentado na Figura 6.2. É possível observar que o atributo raiz da árvore foi determinado pelo atributo IRA. Tal resultado indica que caso o IRA do aluno seja inferior a 2,52, há uma grande probabilidade dele evadir, visto que a quantidade de classificações corretas tiveram uma taxa de 541,11, e de classificações incorretas, apenas 39,18.

É possível também observar que, como o IRA é baseado nas menções dos alunos nas disciplinas, logo caso o aluno reprove muitas disciplinas, seu IRA terá um valor baixo.

Então a primeira conclusão que foi obtida, a partir dessas observações, é que um IRA inferior ou igual a 2,52 é o mais forte indicador de evasão do aluno.

Embora os atributos NASCIMENTO, ANO SAIDA OPCA O e ANO INGRESSO OPCA O tenham aparecidos no resultado, é necessária uma investigação mais profunda para o entendimento do que ocorreu nessas datas.

Continuando com o processo de mineração de dados, o atributo IRA foi retirado para verificar qual seria o segundo atributo que melhor classificaria a evasão do aluno. Ao executar o algoritmo J48 novamente, o resultado obtido é mostrado na Figura 6.3.

```
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
ALGEBRA_1_T1 = SR: Evadiu (136.78/14.0)
ALGEBRA_1_T1 = II: Evadiu (34.19/3.5)
ALGEBRA_1_T1 = MI: Evadiu (26.3/2.69)
ALGEBRA_1_T1 = MM
| INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (93.5/17.37)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (64.17/11.24)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (29.62/5.19)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
| | CALCULO_3_T1 = SR: Evadiu (17.48/2.57)
| | CALCULO_3_T1 = II: Evadiu (11.1/1.63)
| | CALCULO_3_T1 = MI: Evadiu (10.83/3.06)
| | CALCULO_3_T1 = MM
| | | NASCIMENTO <= 1992
| | | | ANO_SAIDA_OPCAO <= 1988: Evadiu (11.95/0.61)
| | | | ANO_SAIDA_OPCAO > 1988
| | | | | CALCULO_2_T1 = SR: Evadiu (9.7/1.77)
| | | | | CALCULO_2_T1 = II: Evadiu (3.49/0.64)
| | | | | CALCULO_2_T1 = MI: Evadiu (3.15/0.58)
| | | | | CALCULO_2_T1 = MM
| | | | | CALCULO_1_T1 = SR: Evadiu (11.48/2.53)
| | | | | CALCULO_1_T1 = II: Evadiu (3.94/0.87)
| | | | | CALCULO_1_T1 = MI
| | | | | | SEXO = M: Evadiu (2.76/0.83)
| | | | | | SEXO = F: Formou (2.87/0.46)
| | | | | CALCULO_1_T1 = MM: Formou (75.21/18.17)
| | | | | CALCULO_1_T1 = MS
| | | | | | RACA = Nao_Dispose_de_Informacao: Formou (28.78/6.98)
| | | | | | RACA = Preta: Evadiu (0.17/0.08)
| | | | | | RACA = Branca: Evadiu (2.02/0.26)
| | | | | | RACA = Parda: Formou (0.44/0.16)
| | | | | | RACA = Amarela: Evadiu (0.15/0.03)
| | | | | | RACA = Nao_Cadastrada: Evadiu (0.12)
| | | | | | RACA = Indigena: Formou (0.0)
| | | | | CALCULO_1_T1 = SS: Formou (4.02/1.14)
| | | | | CALCULO_2_T1 = MS: Formou (45.55/17.5)
| | | | | CALCULO_2_T1 = SS: Evadiu (0.6/0.29)
| | | | | NASCIMENTO > 1992: Evadiu (10.97/0.18)
| | | | | CALCULO_3_T1 = MS: Formou (47.9/22.06)
| | | | | CALCULO_3_T1 = SS: Formou (8.97/3.18)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MS
| | INTRODUCAO_A_CIENTIFICAS_T1 = SR: Evadiu (12.14/2.12)
| | INTRODUCAO_A_CIENTIFICAS_T1 = II: Evadiu (4.32/0.75)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MI: Evadiu (2.05/0.36)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MM: Evadiu (32.47/15.27)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MS
| | | RACA = Nao_Dispose_de_Informacao: Formou (50.57/14.74)
| | | RACA = Preta: Formou (0.26/0.05)
| | | RACA = Branca: Evadiu (3.92/0.95)
| | | RACA = Parda: Evadiu (1.38/0.52)
| | | RACA = Amarela: Evadiu (0.15/0.05)
| | | RACA = Nao_Cadastrada: Evadiu (0.13)
| | | RACA = Indigena: Formou (0.0)
| | INTRODUCAO_A_CIENTIFICAS_T1 = SS: Evadiu (16.56/8.26)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = SS: Evadiu (37.94/16.72)

ALGEBRA_1_T1 = MS
| INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (37.3/5.95)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (26.64/4.25)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (12.3/1.96)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
| | CALCULO_3_T1 = SR: Evadiu (7.23/1.15)
| | CALCULO_3_T1 = II: Evadiu (4.01/0.64)
| | CALCULO_3_T1 = MI: Evadiu (3.84/0.61)
| | CALCULO_3_T1 = MM
| | | VARIABEL_COMPLEXA_1_T1 = SR: Evadiu (4.73/1.7)
| | | VARIABEL_COMPLEXA_1_T1 = II: Evadiu (3.23/0.48)
| | | VARIABEL_COMPLEXA_1_T1 = MI: Evadiu (2.77/0.64)
| | | VARIABEL_COMPLEXA_1_T1 = MM: Formou (46.81/19.2)
| | | VARIABEL_COMPLEXA_1_T1 = MS: Formou (8.96/3.64)
| | | VARIABEL_COMPLEXA_1_T1 = SS: Formou (1.02/0.47)
| | CALCULO_3_T1 = MS
| | | FISICA_1_T1 = SR: Evadiu (1.35/0.23)
| | | FISICA_1_T1 = II: Evadiu (1.06/0.19)
| | | FISICA_1_T1 = MI: Evadiu (1.32/0.23)
| | | FISICA_1_T1 = MM: Evadiu (7.47/2.53)
| | | FISICA_1_T1 = MS: Formou (13.41/4.08)
| | | FISICA_1_T1 = SS: Evadiu (0.12/0.02)
| | CALCULO_3_T1 = SS: Formou (2.93/1.08)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = MS
| | INTRODUCAO_A_CIENTIFICAS_T1 = SR: Evadiu (5.5/0.9)
| | INTRODUCAO_A_CIENTIFICAS_T1 = II: Evadiu (2.2/0.36)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MI: Evadiu (1.1/0.18)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MM: Formou (17.05/6.85)
| | INTRODUCAO_A_CIENTIFICAS_T1 = MS: Formou (28.51/6.34)
| | INTRODUCAO_A_CIENTIFICAS_T1 = SS: Formou (17.36/5.03)
| INTRODUCAO_ALGEBRA_LINEAR_T1 = SS: Formou (28.49/11.48)
ALGEBRA_1_T1 = SS
| CALCULO_1_T1 = SR: Evadiu (10.73/1.46)
| CALCULO_1_T1 = II: Evadiu (5.18/0.71)
| CALCULO_1_T1 = MI: Evadiu (3.86/0.53)
| CALCULO_1_T1 = MM: Evadiu (19.34/6.71)
| CALCULO_1_T1 = MS: Formou (28.77/11.78)
| CALCULO_1_T1 = SS: Formou (16.3/6.08)

Number of Leaves : 76
Size of the tree : 93

Time taken to build model: 0.42 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 928 75.3859 %
Incorrectly Classified Instances 303 24.6141 %
Kappa statistic 0.4166
Mean absolute error 0.3771
Root mean squared error 0.4107
Relative absolute error 80.5486 %
Root relative squared error 84.8971 %
Total Number of Instances 1231
```

Figura 6.3: Resultado obtido na mineração de dados com o algoritmo J48, sem os atributos MATRICALUNO, FOR SAIDA OPCA O e IRA.

É possível observar, de acordo com o resultado mostrado na Figura 6.3, que os atributos NASCIMENTO, RACA, SEXO e ANO SAIDA OPCAOTORNARAM-SE DESINTERESSANTES para a análise, pois exigiria um grau maior de investigação para o entendimento do que ocorreu em cada data e/ou situação específica. Por exemplo, o resultado do classificador mostrou uma condição para os alunos nascidos no ano de 1992, porém não temos informações específicas e suficientemente abrangentes para entender o que ocorreu neste ano. Além disso, esses atributos prejudicaram a análise do resultado, por apresentarem poucas classificações corretas.

Sendo assim, os atributos adicionais foram retirados. Foram mantidos no processo apenas os atributos referentes às tentativas de aprovação nas disciplinas obrigatórias, que apresentam as maiores taxas de reprovação, e o atributo SAIDA. A Figura 6.4 ilustra os atributos que foram mantidos.

28	<input type="checkbox"/>	CALCULO_DE_PROBABILIDADE_1_T1
29	<input type="checkbox"/>	CALCULO_DE_PROBABILIDADE_1_T2
30	<input type="checkbox"/>	CALCULO_DE_PROBABILIDADE_1_T3
31	<input type="checkbox"/>	VARIAVEL_COMPLEXA_1_T1
32	<input type="checkbox"/>	VARIAVEL_COMPLEXA_1_T2
33	<input type="checkbox"/>	VARIAVEL_COMPLEXA_1_T3
34	<input type="checkbox"/>	PROBABILIDADE_E_ESTATISTICA_T1
35	<input type="checkbox"/>	PROBABILIDADE_E_ESTATISTICA_T2
36	<input type="checkbox"/>	PROBABILIDADE_E_ESTATISTICA_T3
37	<input type="checkbox"/>	TEORIA_DOS_NUMEROS_T1
38	<input type="checkbox"/>	TEORIA_DOS_NUMEROS_T2
39	<input type="checkbox"/>	TEORIA_DOS_NUMEROS_T3
40	<input type="checkbox"/>	SAIDA

Figura 6.4: Atributos considerados na mineração de dados com o algoritmo J48, após a remoção dos atributos adicionais.

Ao considerar somente os atributos de tentativas de aprovação nas disciplinas obrigatórias e o atributo SAIDA, o resultado mostrado na Figura 6.5 é já é suficiente para uma análise mais detalhada das disciplinas obrigatórias que mais reprovam em relação ao problema de evasão no curso de Licenciatura em Matemática (Diurno) da UnB.

É importante observar que, o classificador conseguiu classificar corretamente 932 alunos (75,71%), de um total de 1.231 alunos. Apenas 299 alunos (24,28%) fugiram à regra, e foram classificados incorretamente pelo modelo de classificação do resultado.

```

Test mode:      10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
ALGEBRA_1_T1 = SR: Evadiu (136.78/14.0)
ALGEBRA_1_T1 = II: Evadiu (34.19/3.5)
ALGEBRA_1_T1 = MI: Evadiu (26.3/2.69)
ALGEBRA_1_T1 = MM
|  INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (93.5/17.37)
|  INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (64.17/11.24)
|  INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (29.62/5.19)
|  INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
|  |  CALCULO_3_T1 = SR: Evadiu (17.48/2.57)
|  |  CALCULO_3_T1 = II: Evadiu (11.1/1.63)
|  |  CALCULO_3_T1 = MI: Evadiu (10.83/3.06)
|  |  CALCULO_3_T1 = MM
|  |  |  CALCULO_2_T1 = SR: Evadiu (11.43/1.9)
|  |  |  CALCULO_2_T1 = II: Evadiu (5.61/0.93)
|  |  |  CALCULO_2_T1 = MI: Evadiu (4.63/0.77)
|  |  |  CALCULO_2_T1 = MM
|  |  |  |  CALCULO_1_T1 = SR: Evadiu (13.36/2.7)
|  |  |  |  CALCULO_1_T1 = II: Evadiu (5.19/1.05)
|  |  |  |  CALCULO_1_T1 = MI: Formou (6.37/3.08)
|  |  |  |  CALCULO_1_T1 = MM: Formou (79.39/22.89)
|  |  |  |  CALCULO_1_T1 = MS: Formou (35.27/12.68)
|  |  |  |  CALCULO_1_T1 = SS: Formou (5.32/2.25)
|  |  |  |  CALCULO_2_T1 = MS: Formou (50.19/22.13)
|  |  |  |  CALCULO_2_T1 = SS: Evadiu (0.63/0.28)
|  |  |  |  CALCULO_3_T1 = MS: Formou (47.9/22.06)
|  |  |  |  CALCULO_3_T1 = SS: Formou (8.97/3.18)
|  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = MS
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = SR: Evadiu (12.14/2.12)
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = II: Evadiu (4.32/0.75)
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MI: Evadiu (2.05/0.36)
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MM: Evadiu (32.47/15.27)
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MS: Formou (56.41/18.85)
|  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = SS: Evadiu (16.56/8.26)
|  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = SS: Evadiu (37.94/16.72)
|  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (37.3/5.95)
|  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (26.64/4.25)
|  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (12.3/1.96)
|  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
|  |  |  |  |  |  CALCULO_3_T1 = SR: Evadiu (7.23/1.15)
|  |  |  |  |  |  CALCULO_3_T1 = II: Evadiu (4.01/0.64)
|  |  |  |  |  |  CALCULO_3_T1 = MI: Evadiu (3.84/0.61)
|  |  |  |  |  |  CALCULO_3_T1 = MM
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = SR: Evadiu (4.73/1.7)
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = II: Evadiu (3.23/0.48)
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = MI: Evadiu (2.77/0.64)
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = MM: Formou (46.81/19.23)
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = MS: Formou (8.96/3.64)
|  |  |  |  |  |  |  VARIABEL_COMPLEXA_1_T1 = SS: Formou (1.02/0.47)
|  |  |  |  |  |  CALCULO_3_T1 = MS
|  |  |  |  |  |  |  FISICA_1_T1 = SR: Evadiu (1.35/0.23)
|  |  |  |  |  |  |  FISICA_1_T1 = II: Evadiu (1.06/0.19)
|  |  |  |  |  |  |  FISICA_1_T1 = MI: Evadiu (1.32/0.23)
|  |  |  |  |  |  |  FISICA_1_T1 = MM: Evadiu (7.47/2.53)
|  |  |  |  |  |  |  FISICA_1_T1 = MS: Formou (13.41/4.08)
|  |  |  |  |  |  |  FISICA_1_T1 = SS: Evadiu (0.12/0.02)
|  |  |  |  |  |  |  CALCULO_3_T1 = SS: Formou (2.93/1.08)
|  |  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = MS
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = SR: Evadiu (5.5/0.9)
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = II: Evadiu (2.2/0.36)
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MI: Evadiu (1.1/0.18)
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MM: Formou (17.05/6.85)
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = MS: Formou (28.51/6.34)
|  |  |  |  |  |  |  INTRODUCAO_A_CIEN_COMPUTACAO_T1 = SS: Formou (17.36/5.03)
|  |  |  |  |  |  |  INTRODUCAO_ALGEBRA_LINEAR_T1 = SS: Formou (28.49/11.48)
|  |  |  |  |  |  ALGEBRA_1_T1 = SS
|  |  |  |  |  |  |  CALCULO_1_T1 = SR: Evadiu (10.73/1.46)
|  |  |  |  |  |  |  CALCULO_1_T1 = II: Evadiu (5.18/0.71)
|  |  |  |  |  |  |  CALCULO_1_T1 = MI: Evadiu (3.86/0.53)
|  |  |  |  |  |  |  CALCULO_1_T1 = MM: Evadiu (19.34/6.71)
|  |  |  |  |  |  |  CALCULO_1_T1 = MS: Formou (28.77/11.78)
|  |  |  |  |  |  |  CALCULO_1_T1 = SS: Formou (16.3/6.08)
|  |  |  |  |  |  Number of Leaves :      61
|  |  |  |  |  |  Size of the tree :      73
|  |  |  |  |  |  Correctly Classified Instances      932      75.7108 %
|  |  |  |  |  |  Incorrectly Classified Instances      299      24.2892 %

```

Figura 6.5: Resultado do algoritmo J48, ao utilizar os atributos mostrados na Figura 6.4.

## 6.2 Análise dos Resultados

Para analisar o resultado mostrado na Figura 6.5, deve-se estar atento aos resultados que não possuem um quantidade considerável de classificações corretas, pois uma pequena quantidade de classificações corretas para um caso não é determinante para classificar a evasão. Além disso, apesar da técnica de poda de árvore no algoritmo J48 ter reduzido o sobreajuste, ainda sobraram resultados insignificativos.

Podemos perceber que o classificador considerou somente a primeira tentativa na disciplina. Uma hipótese para esse fenômeno é que a maioria dos alunos cursaram a disciplina pelo menos uma vez.

Analisando a árvore do resultado, podemos ver que Álgebra 1 é a disciplina que está na raiz. Assim, podemos concluir que é a disciplina mais importante para determinar a evasão do aluno no curso. A Figura 6.6 mostra que, ao somar-se as quantidades de reprovações pelas menções SR, II e MI, uma média de 197,27 alunos que reprovaram

Álgebra 1, evadiram. Considerando a quantidade total de classificações corretas pelo modelo de classificação, que foi de 932 alunos, esse resultado mostra que 197,27 alunos equivalem a 21,13% dos alunos que evadiram do curso.

```
J48 pruned tree
-----

ALGEBRA_1_T1 = SR: Evadiu (136.78/14.0)
ALGEBRA_1_T1 = II: Evadiu (34.19/3.5)
ALGEBRA_1_T1 = MI: Evadiu (26.3/2.69)
ALGEBRA_1_T1 = MM
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (93.5/17.37)
```

Figura 6.6: Média de alunos que reprovaram em Álgebra 1 e evadiram do curso, por menção.

O resultado ilustrado na Figura 6.7 mostra que, os alunos que foram aprovados em Álgebra 1 com menção MM, na primeira tentativa, e reprovaram Introdução à Álgebra Linear, também na primeira tentativa, evadiram. Ao somar-se as quantidades de reprovações pelas menções SR, II e MI, uma média de 187,29 alunos que reprovaram Introdução à Álgebra Linear e evadiram do curso, representam 20,09% dos alunos que evadiram do curso, considerando a quantidade total de 932 classificações corretas do resultado.

```
ALGEBRA_1_T1 = MM
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (93.5/17.37)
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (64.17/11.24)
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (29.62/5.19)
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
|   |   CALCULO_3_T1 = SR: Evadiu (17.48/2.57)
```

Figura 6.7: Alunos que reprovaram em Introdução à Álgebra Linear e evadiram do curso, após terem sido aprovados em Álgebra 1 com menção MM.

Prosseguindo com a análise, a Figura 6.8 mostra uma pequena quantidade de alunos que, apesar de terem sido aprovados com menção MM em Álgebra 1 e Introdução à Álgebra Linear, evadiram do curso. Eles representam apenas 4,22% dos alunos que evadiram do curso, considerando a quantidade total de 932 classificações corretas.



```

|   INTRODUCAO_ALGEBRA_LINEAR_T1 = MM
|   |   CALCULO_3_T1 = SR: Evadiu (17.48/2.57)
|   |   CALCULO_3_T1 = II: Evadiu (11.1/1.63)
|   |   CALCULO_3_T1 = MI: Evadiu (10.83/3.06)
|   |   CALCULO_3_T1 = MM

```

Figura 6.8: Alunos que reprovaram em Cálculo 3 e evadiram do curso, após terem sido aprovados em Álgebra 1 e Introdução à Álgebra Linear, com menção MM.

Nesse próximo caso, também temos uma quantidade pequena de alunos que, apesar de terem sido aprovados com menção MM em Álgebra 1, Introdução à Álgebra Linear e Cálculo 3, mas reprovaram Cálculo 2 na primeira tentativa, evadiram. Representando apenas 2,32% do total de classificações corretas para evasão. A Figura 6.9 ilustra o resultado em questão.

```

|   |   CALCULO_3_T1 = MM
|   |   |   CALCULO_2_T1 = SR: Evadiu (11.43/1.9)
|   |   |   CALCULO_2_T1 = II: Evadiu (5.61/0.93)
|   |   |   CALCULO_2_T1 = MI: Evadiu (4.63/0.77)
|   |   |   CALCULO_2_T1 = MM

```

Figura 6.9: Alunos que reprovaram em Cálculo 3 e evadiram do curso, após terem sido aprovados em Álgebra 1 e Introdução à Álgebra Linear, com menção MM.

Para prosseguirmos com as análises, iremos adiante considerar apenas os casos que apresentam resultados significativos.

A Figura 6.10 mostra que, apesar dos alunos terem sido aprovados em Álgebra 1 com menção MS, mas reprovados em Introdução à Álgebra Linear, evadiram do curso. Ao somar-se as quantidades de reprovações pelas menções SR, II e MI, temos uma média de 76,24 alunos, que representam 8,18% dos alunos que evadiram do curso, considerando a quantidade total de 932 classificações corretas do resultado.

```

ALGEBRA_1_T1 = MS
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = SR: Evadiu (37.3/5.95)
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = II: Evadiu (26.64/4.25)
|   INTRODUCAO_ALGEBRA_LINEAR_T1 = MI: Evadiu (12.3/1.96)

```

Figura 6.10: Alunos que reprovaram em Introdução à Álgebra Linear e evadiram do curso, após terem sido aprovados em Álgebra 1 com menção MS.

Conforme os resultados apresentados, ao executar o processo de mineração de dados utilizando a técnica de classificação, podemos concluir que as disciplinas que mais estão relacionadas à evasão no curso de Licenciatura em Matemática (Diurno) da UnB, são mostradas na Tabela 6.1.

Tabela 6.1: Disciplinas obrigatórias que mais estão relacionadas à evasão no curso.

Semestre	Código	Nome
5	113107	Álgebra 1
2	113093	Intr. à Álgebra Linear
3	113051	Cálculo 3
2	113042	Cálculo 2

# Capítulo 7

## Conclusões e Trabalhos Futuros

Neste trabalho, obtivemos informações relevantes para a descoberta de disciplinas obrigatórias do curso de Licenciatura em Matemática (Diurno) da UnB que mais estão relacionadas com a evasão de alunos. Com os dados de todos alunos e todas as menções das disciplinas que cursaram no período de 2º/1977 a 1º/2016, foram feitas inicialmente análises estatísticas para a descoberta das disciplinas que mais reprovam. Em seguida, experimentos que utilizam a mineração de dados foram realizados com o objetivo de descobrir quais dessas disciplinas obrigatórias estão mais relacionadas à evasão de alunos.

Como foi observado na mineração de dados, realizada a partir das estatísticas de reprovação nas disciplinas obrigatórias, as disciplinas que mais reprovam e que estão relacionadas com a evasão dos alunos do curso de Licenciatura em Matemática (Diurno) são:

- Álgebra 1
- Introdução à Álgebra Linear
- Cálculo 3
- Cálculo 2

### 7.1 Contribuições

As análises estatísticas e os resultados das minerações de dados dos alunos realizados neste trabalho contribuem para o entendimento da alta taxa de evasão dos alunos do curso de Licenciatura em Matemática (Diurno) da UnB. Este trabalho pode fornecer informações relevantes ao Departamento de Matemática da UnB, especificamente ao curso de Licenciatura em Matemática (Diurno), a fim de auxiliar na criação de medidas para a redução dos índices de evasão no curso.

## 7.2 Trabalhos Futuros

Este trabalho serve como base para estudos mais aprofundados sobre a evasão no curso de Licenciatura em Matemática (Diurno) da UnB, e pode ser seguido por outras ações como:

- Determinar o perfil de um aluno em risco de evasão;
- Analisar com mais rigor e detalhes as disciplinas obrigatórias relacionadas à evasão;
- Calcular quantitativamente o risco de evasão dos alunos que estão atualmente cursando;
- Desenvolver um sistema para rápida e fácil detecção de um aluno em risco de evasão.

# Referências

- [1] Brasil (1996). Lei n.9.394, de 20 de dezembro de 1996: Estabelece as diretrizes e bases da educação nacional. *Constituição da República Federativa do Brasil de 1988, art. 21, art. 3º, art. 4º, art. 43*, 1996. 1, 2
- [2] A. R. J. G. S. Andreoli. Projeto de pesquisa: A evasão na UNIPAMPA: Diagnosticando processos, acompanhando trajetórias e itinerários de formação. *Projeto de Pesquisa*, 2011. 14
- [3] M. T. Angeloni. Elementos intervenientes na tomada de decisão. *Ciência da Informação*, 32:17 – 22, 04 2003. 19
- [4] C. Baggi e D. Lopes. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16:355 – 374, 07 2011. 2
- [5] E. F. S. M. Bonfim. Avaliação do rendimento e evasão de alunos cotistas e não cotistas da Universidade de Brasília. *Monografia: Universidade de Brasília, Brasília*, 2014. 17
- [6] C. B. Cardoso. Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão. *Dissertação (Mestrado em Educação) - Universidade de Brasília, Brasília*, 2008. 18
- [7] A. O. Casartelli, A. C. S. Benso, M. C. Morosini, e R. M. Gessinger. Um estudo sobre os motivos e fatores relacionados com o abandono estudantil na PUCRS. *Libro de Actas II CLABES*, 2012. 15
- [8] J. Castles. Persistence and the adult learner factors affecting persistence in open university students. *Active learning in higher education*, 5(2):166–179, 2004. 6
- [9] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, e R. Wirth. CRISP-DM 1.0 step-by-step data mining guide. Technical report, The CRISP-DM consortium, August 2000. ix, 22, 23
- [10] L. Cheung e A. Kan. Evaluation of factors related to student performance in a distance-learning business communication course. *Journal of Education for Business*, 77(5):257–263, 2002. 5
- [11] L. Chies, G. Graziosi, e F. Pauli. Job opportunities and academic dropout: The case of the University of Trieste. *Procedia Economics and Finance*, 17:63 – 70, 2014. 13

- [12] Portal da Universidade de Brasília. Curso de Matemática. Disponível em: [http://www.unb.br/aluno\\_de\\_graduacao/cursos/matematica](http://www.unb.br/aluno_de_graduacao/cursos/matematica), acesso em 12 de Abril de 2016, 2016. 29, 41
- [13] T. H. Davenport e L. Prusak. *Information Ecology: Mastering the Information and Knowledge Environment*. Oxford University Press, 1st edition, 1997. xii, 19, 20
- [14] Decanato de Ensino de Graduação. Análise do sistema de cotas para negros da Universidade de Brasília. Disponível em: [http://unb2.unb.br/administracao/decanatos/deg/downloads/index/realtorio\\_sistema\\_cotas.pdf](http://unb2.unb.br/administracao/decanatos/deg/downloads/index/realtorio_sistema_cotas.pdf), acesso em 14 de Novembro de 2016, 2013. ix, 16, 17, 40
- [15] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Ministério da Educação (MEC). Resumo técnico do censo da educação superior de 2013. Disponível em: [http://download.inep.gov.br/download/superior/censo/2013/resumo\\_tecnico\\_censo\\_educacao\\_superior\\_2013.pdf](http://download.inep.gov.br/download/superior/censo/2013/resumo_tecnico_censo_educacao_superior_2013.pdf), acesso em 20 de Maio de 2016, 2013. ix, xii, 2, 8, 9, 10, 11, 12
- [16] Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Disponível em: <http://www.dominiopublico.gov.br/download/texto/me001613.pdf>, acesso em 12 de Dezembro de 2015, 1996. 7
- [17] Matrícula Web. Portal de Matrícula da Universidade de Brasília. Curso de Matemática. Disponível em: <https://matriculaweb.unb.br>, acesso em 15 de Abril de 2016, 2016. ix, xii, 30, 31, 32
- [18] C. Donalek. Supervised and unsupervised learning. Disponível em: [http://www.astro.caltech.edu/~george/aybi199/Donalek\\_Classif.pdf](http://www.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf), acesso em 30 de Novembro de 2016, 2011. 25
- [19] P. Dupin-Bryant. Pre-entry variables related to retention in online distance education. *The American Journal of Distance Education*, 18(4):199–206, 2004. 6
- [20] U. M. Fayyad, G. Piatetsky-Shapiro, e P. Smyth. Advances in knowledge discovery and data mining. chapter From Data Mining to Knowledge Discovery: An Overview, pages 1–34. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996. ix, 21, 22
- [21] M. D. Fialho e E. M. T. Prestes. Evasão escolar no curso de pedagogia da UFPB: na compreensão dos gestores educacionais. *Revista do Mestrado Profissional Gestão em Organizações Aprendentes*, 3(1):42–63, 2014. 15
- [22] C. Finnegan, L. Morris, e K. Lee. Differences by course discipline on student behavior, persistence, and achievement in online courses of undergraduate general education. *Journal of College Student Retention: Research, Theory & Practice*, 10(1):39–54, 2008. 7
- [23] F. G. Florencio. Evasão no curso de Licenciatura em Computação da Universidade de Brasília. *Monografia: Universidade de Brasília, Brasília*, 2015. 18

- [24] J. Frydenberg. Persistence in university continuing education online classes. *The International Review of Research in Open and Distributed Learning*, 8(3), 2007. 6
- [25] R. M. Gessinger, M. C. Moraes, L. L. Leite, e V. M. R. Lima. O uso pedagógico de recursos tecnológicos como estratégia para qualificar o ensino e contribuir para a redução da evasão na educação superior. *CLABES, III. Anais.*, 10, 2013. 15
- [26] P. Gordon. *Data Mining: Predicting Tipping Points*, volume 1. Blue Matrix Publications, Paris, FR, 1 edition, 2 2013. 19, 20
- [27] J. Han e M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc. 3rd edition, San Francisco, CA, USA, 2012. ix, 21, 25, 26
- [28] N. Ivankova e S. Stick. Student's persistence in a distributed doctoral program in educational leadership in higher education: A mixed methods study. *Research in Higher Education*, 48(1):93–135, 2007. 6
- [29] W. Kemp. Persistence of adult learners in distance education. *Journal of Education for Business*, 16(2):65–81, 2002. 5
- [30] Y. Lee e J. Choi. A review of online course dropout research: implications for practice and future research. *Educational Technology Research and Development*, 59(5):593–618, 2011. 5
- [31] Y. Levy. Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48(1):185–204, 2007. 6
- [32] A. O. Lima e B. A. C. Costa. Evasão na Universidade de Brasília: um estudo sobre os cursos formadores de docentes para matérias básicas do vestibular, 2012. viii, 95 f. Monografia (Bacharelado em Estatística) — Universidade de Brasília - Brasília. 5, 17, 18
- [33] E. Lima e L. Machado. A evasão discente nos cursos de licenciatura da UFMG. *Educação Unisinos*, 18(2):121–129, 2014. 15
- [34] M. B. D. C. Lobo. Panorama da evasão do ensino superior brasileiro: Aspectos gerais das causas e soluções. pages 01 – 23, 2011. 4, 7
- [35] D. P. M. Miranda. Gestão da evasão nas instituições de ensino superior privado: Um estudo sobre cursos de administração no estado do Espírito Santo. Monografia (Mestrado), Universidade Federal Rural do Rio de Janeiro, 2006. 7
- [36] K. Moore, J. Bartkovich, M. Fetzner, e S. Ison. Success in cyberspace: Student retention in online courses. *Journal of Applied Research in the Community College*, 10(2):107–118, 2003. 5
- [37] C. Morgan e M. Tam. Unravelling the complexities of distance education student attrition. *Distance education*, 20(1):96–108, 1999. 5
- [38] L. Morris, C. Finnegan, e S. Wu. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3):221–231, 2005. 6

- [39] D. L. Olson e D. Delen. *Advanced Data Mining Techniques*. Springer Publishing Company, Incorporated, 1st edition, 2008. ix, 22, 23, 24
- [40] T. Pang-Ning, M. Steinbach, e V. Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005. 20
- [41] L. Paura e I. Arhipova. Cause analysis of students' dropout rate in higher education study program. *Procedia - Social and Behavioral Sciences*, 109:1282 – 1286, 2014. 13
- [42] B. Perry, J. Boman, W. D. Care, M. Edwards, e C. Park. Why do students withdraw from online graduate nursing and health studies education? *Journal of Educators Online*, 5(1):102–153, 2008. 7
- [43] C. Pierrakeas, M. Xeno, C. Panagiotakopoulos, e D. Vergidis. A comparative study of dropout rates and causes for two different distance education courses. *The International Review of Research in Open and Distributed Learning*, 5(2), 2004. 6
- [44] E. Pigliapoco e A. Bogliolo. The effects of psychological sense of community in online and face-to-face academic courses. *International Journal of Emerging Technologies in Learning (iJET)*, 3(4), 2008. 6
- [45] S. S. Prietch e T. A. Pazeto. Estudo sobre a evasão em um curso de licenciatura em informática e considerações para melhorias. in: Workshop de Educação em Informática Bahia-Alagoas-Sergipe (WEIBASE)/ ERBASE. *Anais da 10ª Escola Regional de Computação Bahia-Alagoas-Sergipe (ERBASE)*, 2010. 14
- [46] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993. 26
- [47] P. K. Santos. Evasão na educação superior: uma análise a partir de publicações na ANPED e CAPES (2000 a 2012). *Tercera Conferencia Latinoamericana sobre el Abandono en la Educación Superior (III CLABES)*, 2014. 15
- [48] N. Shin e J. Kim. An exploration of learner progress and drop-out in Korea National Open University. *Distance education*, 20(1):81–95, 1999. 5
- [49] R. L. L. Silva Filho, P. R. Motejunas, O. Hipólito, e M. B. C. M. Lobo. A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, 37:641 – 659, 12 2007. 4
- [50] S. Tello. An analysis of student persistence in online education. *International Journal of Information and Communication Technology Education*, 3(3):47–62, 2007. 6
- [51] L. M. V. Tigrinho. Evasão escolar nas instituições de ensino superior, 2008. *Revista Gestão Universitária da América Latina – GUAL*, pages 173:01–14, 2011. 12
- [52] E. D. Velez. America's college drop-out epidemic: Understanding the college drop-out population. Disponível em: <http://www.air.org/sites/default/files/downloads/report/AIR-CALDER-Understanding-the-College-Dropout-Population-Jan14.pdf>, acesso em 02 de Novembro de 2016, 2014. 12



- [53] P. Willging e S. Johnson. Factors that influence student's decision to dropout of online courses. *Journal of Asynchronous Learning Networks*, 13(3):115–127, 2009. 6
- [54] I. H. Witten e E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005. 21
- [55] A. Woodley, P. Lange, e G. Tanewski. Student progress in distance education: Kemmer's model re-visited. *Open Learning*, 16(2):113–131, 2001. 5
- [56] X. Wu, V. Kumar, J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. Zhou, M. Steinbach, D. J. Hand, e D. Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1):1–37, December 2007. 26
- [57] M. Xenos, C. Pierrakeas, e P. Pintelas. A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. *Computers & Education*, 39(4):361–377, 2002. 5

# Anexo I

## Formato Completo do Banco de Dados

@RELATION ALUNO

@ATTRIBUTE MATRICALUNO STRING

@ATTRIBUTE SEXO {M,F}

@ATTRIBUTE NASCIMENTO NUMERIC

@ATTRIBUTE UF {SP,MG,DF,PE,GO,MA,RN,RJ,BA,RS,CE,MS,  
PB,PI,ES,SC,PA,PR,TO,AM,AL,MT,RR,SE,RO,AC}

@ATTRIBUTE TIPO\_ESCOLA {Nao\_Declarado,Publica,Particular}

@ATTRIBUTE RACA {Nao\_Dispos\_de\_Informacao,Preta\_,  
Branca,Parda,Amarela,Nao\_Cadastrada,Indigena}

@ATTRIBUTE ANO\_INGRESSO\_OPCAO NUMERIC

@ATTRIBUTE SEMESTRE\_INGRESSO\_OPCAO NUMERIC

@ATTRIBUTE FORMA\_INGRESSO\_OPCAO {Vestibular,  
Mudanca\_de\_Habilitacao,Dupla\_Habilitacao,Portador\_Diploma\_Curso\_Superior,  
Convenio\_Int,Transferencia\_Obrigatoria,Novo\_Vestibular,Mudanca\_de\_Curso,  
Duplo\_Curso,Transferencia\_Facultativa,PAS,Convenio\_Andifes,  
Registro\_de\_Diploma,ENEM,SISU}

@ATTRIBUTE FOR\_SAIDA\_OPCAO {Mudanca\_de\_Curso,Vestibular\_p/outra\_Habilitacao,  
Desligamento\_Rendimento\_Academico,Formatura,Transferencia,Desligamento\_Abandono,  
Desligamento\_Voluntario,Deslig\_Nao\_Cumpriu\_condicao,Mudanca\_de\_Habilitacao,  
Falecimento,Rep\_3\_vezes\_na\_mesma\_Disc\_Obrig,  
Novo\_Vestibular,Anulacao\_de\_Registro,Esta\_Cursando}

@ATTRIBUTE ANO\_SAIDA\_OPCAO NUMERIC

@ATTRIBUTE SEMESTRE\_SAIDA\_OPCAO NUMERIC

@ATTRIBUTE IRA NUMERIC

@ATTRIBUTE FISICA\_1\_T1 {SR,II,MI,MM,MS,SS}

@ATTRIBUTE FISICA\_1\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_1\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_1\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_1\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_1\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_3\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_3\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_3\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_ALGEBRA\_LINEAR\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_ALGEBRA\_LINEAR\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_ALGEBRA\_LINEAR\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_A\_CIEN\_COMPUTACAO\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_A\_CIEN\_COMPUTACAO\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE INTRODUCAO\_A\_CIEN\_COMPUTACAO\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_2\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_2\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE CALCULO\_2\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ALGEBRA\_1\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ALGEBRA\_1\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ALGEBRA\_1\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_2\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_2\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_2\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_1\_EXPERIMENTAL\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_1\_EXPERIMENTAL\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE FISICA\_1\_EXPERIMENTAL\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ANALISE\_1\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ANALISE\_1\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE ANALISE\_1\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE VARIABEL\_COMPLEXA\_1\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE VARIABEL\_COMPLEXA\_1\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE VARIABEL\_COMPLEXA\_1\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE TEORIA\_DOS\_NUMEROS\_T1 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE TEORIA\_DOS\_NUMEROS\_T2 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE TEORIA\_DOS\_NUMEROS\_T3 {SR,II,MI,MM,MS,SS}  
 @ATTRIBUTE SAIDA {Formou,Evadiu}

@DATA

% A seguir, todos dados dos alunos referentes aos atributos  
% Os dados faltosos estão representados pelo símbolo "?"

aac1e1aad3a2063f09bfb9c998850a27d,M,1957,SP,Nao\_Declarado,Nao\_Dispose\_de\_Informacao,  
1977,2,Vestibular,Mudanca\_de\_Curso,1984,2,?,?,?,?,MM,SR,SR,?,?,?,?,?,?,?,?,?,  
?,?,?,?,?,?,?,?,?,?,?,?,?,Evadiu

...

4f6a8750a03035fe27a66bfdc616f154,M,1998,DF,Particular,Parda,2016,1,SISU,  
Novo\_Vestibular,2016,1,0.0,?,?,?,?,MS,?,?,?,?,?,?,?,?,MM,?,?,?,?,?,?,?,?,  
?,?,?,?,?,?,?,?,?,?,?,?,?,Evadiu

# Anexo II

## Currículo Vigente em 1º/2005

Tabela II.1: Currículo vigente em 1º/2005.

Semestre	Código	Nome
1	113034	Cálculo 1
1	113913	Introdução à Ciência da Computação
2	113042	Cálculo 2
2	118001	Física 1
2	118010	Física 1 Experimental
2	113093	Introdução à Álgebra Linear
2	115045	Probabilidade e Estatística
3	113051	Cálculo 3
3	113301	Equações Diferenciais
3	118028	Física 2
3	118036	Física 2 Experimental
3	124966	Fund. do Desenvolvimento e Aprendizagem
4	113417	Cálculo Numérico
4	117161	Geometria 1
4	191027	Psicologia da Educação
4	113115	Teoria dos Números
5	113611	Álgebra para o Ensino 1 e 2
5	113107	Álgebra 1
5	117170	Geometria 2
6	113204	Análise 1
6	192015	Didática Fundamental
6	194221	Organização da Educação Brasileira
7	113069	Variável Complexa 1

# Anexo III

## Algoritmo Utilizado para Limpeza dos Dados

```
#!/usr/bin/python
# -*- coding: utf-8 -*-

import sys
import md5
from unicodedata import normalize

DEBUG = True #False ou True

class arquivoCSV(object):
    """Classe para tratar o arquivo CSV 2016"""

    # Indices dos atributos a serem mantidos
    manter = [2, 3, 5, 12, 25, 29, 34, 36, 37, 38, 39, 40, 41, 42, 44, 45,
              46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 63, 64,
              65, 66, 67, 68, 69, 71, 72]

    def __init__(self, arquivo):
        try:
            self.arquivo = open(arquivo, 'rb')
        except IOError, e:
            print >> sys.stderr, e
            sys.exit()
        self.__readAllFile()

    def __del__(self):
```

```

        self.arquivo.close()

def __readLine(self):
    try:
        line = self.arquivo.readline()
    except IOError, e:
        print >> sys.stderr, e
        sys.exit()

    line = line.rstrip('\r\n')
    line = line.decode('unicode_escape')
    line = self.normalizarNome(line)
    line = line.split(';')
    # Remove aspas duplas
    for l in range(len(line)):
        if line[l].startswith('"') and line[l].endswith('"'):
            line[l] = line[l][1:-1]
    return line

def __readHeader(self):
    self.header = []
    for attr in self.__readLine():
        if attr.upper() == 'ANO(DATA)':
            attr = 'ANO_DATA'
        if not attr.upper().find('CODDISC'):
            self.ano_referencia = attr.upper().split('CODDISC')[1]
            attr = 'COD_DISCIPLINA'
            self.header.append(attr.upper())
            self.header.append('DISCIPLINA_OBR')
            continue
        if not attr.upper().find('MENCAO'):
            attr = 'MENCAO'
        if not attr.upper().find('NOME_DISCIPLINA'):
            attr = 'NOME_DISCIPLINA'
        self.header.append(attr.upper())
    self.header.append('ANO_SEM_REFERENCIA')
    self.__removeAttributes(self.header)

def __readAlunos(self):
    self.alunos = []

```

```

while True:
    line = self.__readLine()
    if line != ['']:
        self.__removeAttributes(line)
        self.__hideSensitive(line)
        self.alunos.append(line)
    else:
        break

def __removeAttributes(self, source):
    c = 0
    for i in self.remover:
        index = int(i - c)
        if DEBUG:
            print 'Removendo atributo: ' + source[index] + '[' + str(i) + ']'
        del source[index]
        c += 1

def __hideSensitive(self, source):
    matricula = md5.new()
    matricula.update(source[0])
    source[0] = matricula.hexdigest()
    source[2] = source[2].split('/')[2]
    if source[1].upper() == 'Masculino'.upper():
        source[1] = 'M'
    else:
        source[1] = 'F'
    pass

def __readAllFile(self):
    self.arquivo.seek(0)
    self.__readHeader()
    self.__readAlunos()

def normalizarNome(self, source):
    source = source.rstrip()
    source = source.replace('-', '_')
    source = source.replace(',', '_')
    source = source.replace(' ', '_')
    source = source.replace('.', '_')

```



```

        source = source.replace('___','_')
        source = source.replace('__','_')
        return normalize('NFKD', source).encode('ASCII','ignore')

def getAnoRef(self):
    return self.ano_referencia

def getAlunos(self):
    return self.alunos

def getHeader(self):
    return self.header

def test():
    csv = arquivoCSV(sys.argv[1])
    print csv.getAlunos()
    print csv.getHeader()

    cabecalho = ''
    for c in range(len(csv.getHeader())):
        cabecalho += csv.getHeader()[c] + '[' + str(c) + '], '

    print cabecalho

    print 'Quantidade de atributos: %d' % (len(csv.getHeader()))
    #for l in range(len(csv.getHeader())):
    #    print csv.getHeader()[l] + '[' + str(l) + ']'
    print 'Quantidade de registros: %d' % len(csv.getAlunos())

    print 'Verificando a consistencia (quantidade de atributos) de cada aluno...'

    erro = 0
    for aluno in csv.getAlunos():
        if len(aluno) + 2 != len(csv.getHeader()):
            print 'Um erro ocorreu! len(aluno) != len(csv.getHeader())!'
            print aluno
            erro = 1

    if erro == 0:
        print 'Todos alunos foram verificados com sucesso!'

```

```
    else:
        print 'Erros foram encontrados ao verificar a consistencia de dados dos alunos'

if DEBUG:
    test()
```

## Anexo IV

# Algoritmo Utilizado para Inserção dos Dados no MySQL

```
#!/usr/bin/python
# -*- coding: utf-8 -*-

import sys
import MySQLdb
from arquivo_csv_2016 import arquivoCSV

DEBUG = False #False ou True

csv = arquivoCSV(sys.argv[1])
if DEBUG:
    print csv.getAlunos()
    print csv.getHeader()
    print 'Ano sem referencia: ' + csv.getAnoRef()

db = MySQLdb.connect(host="localhost", user="root", passwd="", db="LICMAT2016")
cur = db.cursor()

cod_disciplinas_obr = [113107,117421,117501,113204,125172,113034,113042,113051,
113824,125156,194531,118001,118010,118028,117161,117170,105881,117471,117480,
113093,113913,194221,117510,117439,113115,113069,117056,124966,192015,
113611,115045,117153]
nome_disc_obr = ['ÁLGEBRA_1','ÁLGEBRA_PARA_O_ENSINO_1','ÁLGEBRA_PARA_O_ENSINO_2',
'ANALISE_1','APRENDIZAGEM_NO_ENSINO','CALCULO_1','CALCULO_2','CALCULO_3',
'CALCULO_DE_PROBABILIDADE_1','DESENVOL_PSILOGICO_E_ENSINO','DIDATICA_FUNDAMENTAL',
'FISICA_1','FISICA_1_EXPERIMENTAL','FISICA_2','GEOMETRIA_1','GEOMETRIA_2',
```

```
'GEOMETRIA_ANALÍTICA_MATEMÁTICA', 'GEOMETRIA_PARA_O_ENSINO_1', 'GEOMETRIA_PARA_O_ENSINO_2',
'INTRODUCAO_A_ALGEBRA_LINEAR', 'INTRODUCAO_A_CIEN_COMPUTACAO',
'ORGAN_DA_EDUCACAO_BRASILEIRA', 'REGÊNCIA_1', 'REGÊNCIA_2', 'TEORIA_DOS_NUMEROS',
'VARIABEL_COMPLEXA_1', 'EST_SUP_DE_REGENCIA_EM_MAT', 'FUND_DESENV_E_APRENDIZAGEM',
'DIDATICA_FUNDAMENTAL', 'ALGEBRA_PARA_ENSINO_1_e_2', 'PROBABILIDADE_E_ESTATISTICA',
'ESTAGIO_EM_LABORATORIO_DE_ENSINO_DE_MATEMATICA']
```

```
def gerarInsert(aluno):
    query = 'INSERT INTO ALUNO VALUES (NULL,'
    for i in range(len(aluno)):
        if aluno[i] == '':
            query += 'NULL,'
        else:
            query += '"' + str(aluno[i]) + ','

    if i == 32:
        obr = False
        for cod in cod_disciplinas_obr:
            if int(aluno[i]) == cod:
                obr = True
                break

        if obr == True:
            query += '1,'
        else:
            query += '0,'

    query += '"' + csv.getAnoRef() + '"'
    query += ')'
    return query

print '\n---\nINSERINDO DADOS NO BANCO DE DADOS\n---\n'
print 'Arquivo: ' + sys.argv[1] + '.csv'
```

```
num_aluno = 1
```

```
for aluno in csv.getAlunos():
    try:
        cur.execute(gerarInsert(aluno))
        db.commit()
```

```

        if DEBUG:
            print 'INSERINDO ALUNO: ' + str(num_aluno)
            num_aluno += 1
    except IOError, e:
        db.rollback()
        print 'ERRO!'
        print >> sys.stderr, e
        sys.exit()

print 'ULTIMO INSERT:'
print gerarInsert(csv.getAlunos()[len(csv.getAlunos()) - 1])

```